(in press). *Journal of Experimental Psychology: General*. To appear February 2007.

Representing Causation

Phillip Wolff

Emory University

Please address correspondence to:
Phillip Wolff
Emory University
Department of Psychology
532 N. Kilgo Circle
Atlanta, GA 30322
Office #: 404-727-7140
Fax #: 404-727-0372
pwolff@emory.edu

Abstract

The *dynamics model*, which is based on Talmy's (1988) theory of *force dynamics*, characterizes causation as a pattern of forces and a position vector. In contrast to counterfactual and probabilistic models, the dynamics model naturally distinguishes between different cause-related concepts and explains the induction of causal relationships from single observations. Support for the model is provided in experiments in which participants categorized 3D animations of realistically rendered objects with trajectories that were wholly determined by the force vectors entered into a physics simulator. Experiments 1-3 showed that causal judgments are based on several forces, not just one. Experiment 4 demonstrated that people compute the resultant of forces using a qualitative decision rule. Experiments 5 and 6 showed that a dynamics approach extends to the representation of social causation. Implications for the relationship between causation and time are discussed.

Key Words: causation, causal models, causal learning, knowledge structures, lexical semantics

Representing Causation

Theories of causal representation are a natural starting point for the study of causal cognition. They tell us, for example, what it is that people learn when they induce causal relationships and what it is that they mean when they use causal language. They also place constraints on the kinds of processes that might be used in causal reasoning. The story would be simple, perhaps, if the concept of CAUSE were a conceptual primitive, as some have proposed (e.g., Anscombe, 1971; Carter, 1976; Jackendoff, 1983; Miller & Johnson-Laird, 1976; Norman, Rumelhart, & the LNR Research Group, 1975; Schank, 1972). But work in psychology, linguistics, philosophy and artificial intelligence suggests that causal relationships can be decomposed. Two dominant approaches to decomposition have emerged.

According to *dependency models*, causal relationships are represented as contingencies between causes and effects.[1] Such contingencies are encoded in terms of either counterfactual conditionals or probabilities. Models that represent causation in terms of probabilities in particular will be referred to as *probability distribution models*. The second major approach to causal representation is specified in *physicalist models*. The basic assumption in these theories is that physical causal relationships can be described and represented in terms of physical quantities in the world, such as energy, linear and angular momentum, impact forces, chemical forces, and electrical forces, among others. In these theories, non-physical causation is explained by analogy to physical causation (see Gentner, Holyoak, & Kokinov, 2001).

In general, phyiscalist theories of causation have had limited impact in the cognitive sciences in part because of uncertainty about the units of cognition and relevant physical quantities involved. These limitations are addressed in a new physicalist model of causation, the *dynamics model*, which reduces causation to patterns of forces and a position vector. In a series of six experiments, I show that this model has several advantages over dependency models and other physicalist models of causation. In particular, it has better extensional adequacy, that is, it more accurately picks out the range of situations that are considered causal without including situations that are non-causal. In addition, it explains how various cause-related concepts might be identified on the basis of a single observation or trial.

Experiments 1-3 show that the dynamics model, unlike other models of causation, is able to distinguish the concept of CAUSE from related concepts as well as from non-causal interactions.

[1] Sloman (2006) has referred to these accounts as “make-a-difference” models.

Experiment 4 provides evidence in support for one of the main assumptions of the dynamics model: that people represent causal relationships in terms of forces. Experiments 5 and 6 show that the dynamics model is not limited to physical causation, but can account for the representation of social causation as well. In the next sections I describe dependency and physicalist accounts of causation in greater detail. I conclude that while dependency models capture important aspects of causation, they do not capture the core properties. As such, these models are best viewed of as tests for causation, not accounts of how it is represented in the mind (Bunge, 1959; Bigelow & Pargetter, 1990; Fair, 1979).

Dependency models

In dependency models, the effect is in some way contingent upon the cause. One major way of representing contingencies is in terms of statistical dependencies. A second way is in terms of counterfactual dependencies. In both types of dependency models, the properties of a causal event matter very little to the way the events are represented. In probability accounts in particular, all that is required is that the events be countable. This feature of dependency models makes it difficult for them to distinguish causation from related concepts or to account for the induction of causal relationships from individual observations. Before discussing these limitations, I review the logic motivating several of the most prominent dependency models.

A major subclass of dependency models represents causal relationships in terms of statistical dependencies. A statistical dependency exists when the probability of an effect in the presence of a cause, $P(E|C)$, is greater than the probability of an effect in the absence of a cause, $P(E|{\sim}C)$. In Cheng and Novick's (1991, 1992) *probabilistic contrast model*, this dependency is represented by subtracting the probability of an effect, E, in the presence of a candidate cause, C, from the probability of the effect in the absence of the candidate cause, in other words, $\Delta P = P(E|C) - P(E|\neg C)$. When ΔP differs from zero, it implies that the cause and the effect covary. In contrast to other covariational models, this one holds that ΔP is computed with respect to a "focal set" of events, that is, a subset of the universal set of events. The probabilistic contrast model distinguishes two main types of causal relationships. A generative, or facilitative, cause is implied if the probability of the effect is noticeably greater in the presence of a possible cause than in its absence. For example, the probability of cancer in the presence of smoking is greater than the probability of cancer in the absence of smoking, which leads us to infer that "smoking causes cancer." An inhibitory, or preventative, cause is inferred if the effect occurs noticeably more often in the absence of the causal factor than in its presence. For example, the probability of cancer is lower in the presence of certain antioxidants than in their absence.

Of course, not all correlations are associated with causation. Thunder correlates with power outages, but thunder does not cause power outages. To distinguish causal from non-causal correlations, it is important to control for alternative causes. In effect, we need to evaluate correlations on the basis of focal sets in which the candidate cause of interest does not covary (is not confounded) with alternative causes. This principle is built into Cheng's (1997, 2000) power PC theory. In this model, causal judgments are not based on covariation directly, but rather upon a theoretical entity—causal power—that can be estimated from covariation, provided certain conditions are honored (see Luhmann & Ahn, 2005 for a critical analysis of these assumptions). Specifically, generative causal power is estimated from the normalization of ΔP by $1-P(E|\neg C)$, that is, $p_{cause} = \Delta P / 1 - P(E|\neg C)$ while preventative causal power, $p_{prevent}$, is estimated from the normalization of $-\Delta P$ by $P(E|\neg C)$, that is, $p_{prevent} = -\Delta P / P(E|\neg C)$. Cheng (1997, 2000) shows that the equations associated with generative and preventative causal power apply only when all alternative causes of the effect vary independently of the candidate cause. Thus, the power PC theory can motivate why people prefer to draw causal inferences from focal sets in which alternative causes are not confounded with the candidate cause of interest. Further, the power PC theory can also explain why zero contrasts (i.e., $\Delta P \approx 0$) are sometimes interpreted as non-causal but at other times are uninterpretable. When both $\Delta P \approx 0$ and $p_{cause} = 0$, a zero contrast will be interpreted as non-causal. However, when $\Delta P \approx 0$ and p_{cause} is undefined, as occurs when $P(E|\neg C) = 1$ due to division by 0, the zero contrast is uninterpretable.

The probabilistic contrast and Power PC models focus on how people represent individual causal relationships. Clearly, though, people can also represent and reason about systems of causal relationships. Bayesian network theories have been used to investigate these kinds of representations (Gopnik, Glymour, Sobel, Shultz, Kushnir, & Danks, 2004; Glymour, 2001; Hagmayer & Waldmann, 2000; Hagmayer, Sloman, Lagnado, & Waldmann, in press; Lagnado, Waldmann, Hagmayer, & Sloman, in press; Pearl, 2000; Sobel, Tenenbaum, & Gopnik, 2004; Sloman & Lagnado, 2002; Sloman, 2005; Tenenbaum & Griffiths, 2001). A Bayesian network consists of a set of nodes, corresponding to variables, a set of arcs (arrows) indicating statistical dependencies, and a set of functions that define probability distributions for each node in the network, and by extension, the network as a whole. A very simple Bayesian network might consist of only three variables, e.g., X, Y, Z, linked by two arrows, $X \rightarrow Y \rightarrow Z$. The arrows imply causal relationships. They do so, in part, by indicating how variables are conditionalized upon other variables. For example, in a network such as $X \rightarrow Y \rightarrow Z$, the probability of Z is conditionalized on Y, $P(Z|Y)$, and the probability of Y is conditionalized on X, $P(Y|X)$. In contrast, in a network such as $X \rightarrow Y \leftarrow Z$, the probability Y is conditionalized on both X and Z, $P(Y|X \& Z)$.

However, the representation of causal relationships in Bayesian networks involves more than conditional probabilities. Conditional probabilities do not necessarily imply statistical dependencies. If, for example, $P(Z|Y) = P(Z)$, it would entail that the variables Y and Z were independent. However, in the construction of a Bayesian network, an arrow is created between two variables only if the two variables are statistically dependent. Thus, in the network in which an arrow is drawn between two variables, such as $Y \rightarrow Z$, the assumption is that the value of a child variable depends statistically on the immediate parent, hence $P(Z|Y) \neq P(Z)$. The process by which the arrows are created is very much a matter of current research (Gopnik et al., 2004). Pearl (2000) has proposed a method of intervention for distinguishing between different network structures (see also Sloman & Lagnado, 2002, 2005; Waldmann & Hagmayer, 2005). Others have proposed that the process of learning causal relations can occur through Bayesian inference (Tenenbaum & Griffiths, 2001; Griffiths & Tenenbaum, in press). Yet other researchers have offered a constraint-based solution, whereby algorithms are used to discover which variables covary with one another (see Gopnik et al., 2004). Crucially, in all of these discovery procedures, the causal links are discovered on the basis of statistical dependencies.

The second major way of encoding causal dependencies is in terms of counterfactuals. According to a counterfactual analysis of causation, an event *c* is a cause of an event *e* if and only if it is the case that if *c* had not occurred, *e* would not have occurred (Lewis, 1973; Mackie, 1974; see also Dowty, 1979; Kahneman & Tversky, 1982; Mandel & Lehman, 1996; Spellman & Mandel, 1999; Wells & Gavanski, 1989; Sloman, 2005). Consider a situation in which a boy throws a water balloon at his father, the balloon bursts, and the father gets wet. According to a counterfactual analysis of causation, the situation is causal because if the boy had not thrown the water balloon, his father would not have gotten wet. The counterfactual establishes that the occurrence of the effect depends on the occurrence of the cause.

Challenges for Dependency Models

Probability distribution models and counterfactual models have both provided impressively close estimates of people's judgments of causation (e.g., Buehner & Cheng, 1997; Griffiths & Tenenbaum, in press; Lober & Shanks, 2000; Mandel & Lehman, 1996; Spellman, 1996; Spellman & Mandel, 1999; Wasserman, Elek, Chatlosh, & Baker, 1993). However, a closer examination of these models suggests that the statistical and counterfactual dependencies are better viewed as tests for the presence of causation than as representations of its essential elements.

Problems in extension. Theories of causation should be able to pick out the range of situations that people judge to be causal. However, current dependency models do not categorize causal situations in the same way as people. In particular, dependency models conflate the concepts of CAUSE and ENABLE. For people, these concepts are similar but not synonymous. In most contexts, they are not interchangeable, as illustrated by the sentences in (1) and (2).

(1) a. A cold wind caused him to close the window.
    b. A crank enabled him to close the window.

(2) a. ?A cold wind enabled him to close the window.
    b. ?A crank caused him to close the window.

The sentences in (1a) and (1b) are perfectly acceptable. However, if the verbs in (1a) and (1b) are switched, the resulting sentences (2) sound quite odd. Despite the clear difference between CAUSE and ENABLE, probability distribution models effectively treat these concepts as synonymous (Goldvarg & Johnson-Laird, 2001; Wolff & Song, 2003). In the case of the situations described in (1), the probability of closing the window is greater in the presence of a cold wind than in its absence. But it is also the case that the probability of closing the window is greater in the presence of the crank than in its absence. This implies that, without further assumptions, probability distribution models, such as the probability contrast model and the Power PC model, cannot distinguish causers from enablers since the ΔP for both causers and enablers is positive.[2] The problem of distinguishing causers from enablers extends to Bayesian

[2] Cheng and Novick (1991; Cheng, 1997) have argued that the probabilistic contrast and the Power PC models are able to differentiate causers and enablers through use of focal sets. Specifically, they propose that an enabler is a candidate causal factor that is constantly present in the focal set under consideration (making $P(\text{effect}|\neg\text{cause}$ undefined) but covaries with the effect in other possible focal sets. Wolff and Song (2003) identify several problems with this proposal. Among others, this account implies that prevalent causes, like gravity, should be viewed as enablers since they are likely to be constantly present in at least one of the reasoner's focal set of events, but might

nets. As noted by Sloman (2005), the distinction between CAUSE and ENABLE is not specified in the arrows that connect the variables. Of course, it might be possible to parameterize a Bayesian net such that a particular range of probabilities would be assigned a particular causal interpretation, but such interpretations are not constrained by the causal networks themselves. In addition to conflating the notions of CAUSE and ENABLE, probability distribution models also conflate the notions of PREVENT and DESPITE[3] since, in both of these relationships, the agents serve to lower the probability of their effects.

Counterfactual approaches to the representation of causation fare no better. Consider, again, the sentences in (1). According to a counterfactual criterion, the crank—which is an enabler—should be construed as a cause because it is true that if the crank had not been present, the window would not have closed. However, as shown in (1b), the crank is not easily viewed as a cause.[4] Lombard (1990) suggests that the counterfactual analysis of causation might be saved if only enabling conditions could be filtered out from consideration prior to application of the counterfactual test. However, in such a solution, the filter itself would arguably be the most interesting part of the theory (see also Wolff & Song, 2003; Goldvarg & Johnson-Laird, 2001).

The problems with a counterfactual account of causation extend well beyond the distinction between CAUSE and ENABLE (Mandel, 2003). A well known difficulty of the counterfactual approaches to causation is overdetermination (Spellman, Kincannon, & Stose, 2005; Sloman, 2005). For example, if a son and a daughter simultaneously threw water balloons at their father, most would agree that both were guilty of causing their father to get wet. However, according to a counterfactual criterion, neither should be considered a cause because if the son's balloon had not been thrown, the father would still have gotten wet (because of the daughter's balloon). A counterfactual criterion associates causation with a necessary condition, but people seem to associate causation with sufficient conditions (Mandel, 2003). Further evidence against a counterfactual analysis is reflected in Mandel and Lehman's (1996) finding that people's counterfactual judgments of what might change a particular outcome often do not agree with

---

covary with an effect such as falling in another focal set. However, in contrast to this prediction, it sounds more natural to say *gravity caused the ball to fall* than *gravity enabled the ball to fall*.

[3] The concept of DESPITE is one of the basic concepts implied by the dynamics model.

[4] The failures of the counterfactual criterion are not saved by Mackie's INUS condition: like the cold wind, the crank can be viewed as an *Insufficient* but *Necessary* part of a complex set of factors that together were *Unnecessary* but *Sufficient* for closing the window. As many have noted, the concept of CAUSE, as well as ENABLE, cannot be characterized in terms of necessity or sufficiency (e.g., Cheng & Novick, 1991, 1992; Einhorn & Hogarth, 1986; Hart & Honoré, 1985; Turnbull & Slugoski, 1988; Wolff & Song, 2003; Wolff, Song, & Driscoll, 2002).

their judgments of what caused the outcome (see also Spellman, et al., 2005; Spellman & Mandel, 1999). Perhaps even more embarrassing, the counterfactual criterion sometimes identifies noncausal factors as causal (Wolff & Song, 2003). For example, according to the counterfactual criterion, birth causes death, given the truthfulness of the counterfactual *if birth had not occurred, death would not occur*. There is evidence that counterfactual thinking can influence people's causal reasoning (Spellman et al., 2005; Mandel, 2003), but neither the psychological or philosophical literature support the idea that causal relationships can be reduced to counterfactual conditionals.

Single-instance identification of causal relations. Causal relationships can sometimes be established on the basis of a single observation (e.g., Ahn & Kalish, 2000; Goldvarg & Johnson-Laird, 2001; Luhmann & Ahn, 2005; Wolff & Song, 2003; Tenenbaum & Griffiths, 2003; Sloman, 2005; White, 1999). This ability is demonstrated in the case of inferring the cause of a single collision event (e.g., Hubbard & Ruppel, 2002; Kruschke & Fragassi, 1996; Michotte, 1946/1963; Schlottmann & Shanks, 1992; Scholl & Nakayama, 2004; White, 1999). It is also demonstrated in people's ability to infer the cause of a particular school closing, plane crash, or forest fire (e.g., Kalish & Ahn, 2000; Goldvarg & Johnson-Laird, 2001; Wolff & Song, 2003).

Single-instance identification is highly problematic for probability distribution models since establishing reliable probabilities requires multiple observations (Tenenbaum & Griffiths, 2001). Lien and Cheng (2000) address this issue by proposing that causal events might be recognized from single observations by means of causal categories learned earlier on the basis of covariational information (see also Cheng, 1993). According to this approach, people may not recognize, for example, that a collision event or a particular kind of social event is causal the first time they see it. However, after many exposures, people may begin to notice that collision events and certain social events involve covariation between candidate causes and effects. Once such events are recognized as causal, they could be stored in memory as causal categories so that the next time such events are encountered they can be recognized as causal right away.

Certain aspects of Lien and Cheng's (2000) proposal are most certainly true, but it seems unlikely that our ability to identify causal events should depend solely on storage of a prior regularity (Fair, 1979). Such an account would imply that people should not be able to recognize causal connections that run counter to their expectations. But this is clearly not the case. The passengers on the Titanic had probably never experienced (or even heard of) ice breaking through iron, yet when it sank, they were able to recognize the cause. People can recognize causation even when it defies regularities in their experience (Fair, 1979). Another problem with causal categories concerns their acquisition. According to Lien and Cheng, causal categories are formed on the basis of covariational information, which, as discussed above, is not enough to differentiate CAUSE from ENABLE, or PREVENT from DESPITE. Hence, a causal category approach cannot explain how people differentiate causation from related concepts on the basis of a single observation.

In contrast to probabilistic theories, counterfactual theories of causation can be applied to the identification of causal relations on the basis of a single observation (Lewis, 1973). Nevertheless, counterfactual theories seem to require the very knowledge they are intended to provide (Fair, 1979; Spellman et al., 2005). For example, we might conclude that spoiled eggs caused the woman to get a stomachache, given the acceptability of the counterfactual *if the woman had not eaten the eggs, she would not have gotten a stomachache*. But in order to imagine this possible counterfactual world, we would have to already know that spoiled eggs can cause stomachaches. And if this fact is already known, constructing the counterfactual serves no point.

The problems faced by dependency models suggest that there is more to peoples' representations of causation than statistical or counterfactual dependencies. Specifically, such dependencies may be the consequence of a more grounded and embodied type of knowledge representation. Proposals that have attempted to capture this type of representation can be called physicalist theories.

Physicalist models of causation

The basic idea in physicalist approaches to causation is that such relationships can be reduced to physical quantities in the world, such as energy, momentum, linear and angular momentum, impact forces, chemical forces, and electrical forces, among others. For example, according to Aronson's (1971) Transference Theory, causation implies contact between two objects in which a quantity possessed by the cause (e.g., velocity, momentum, kinetic energy, heat, etc.) is transferred to the effect. Another transference theory is proposed by Fair (1979), who holds that causes are the source of physical quantities, energy, and momentum that flow from the cause to the effect. A key difference between Aronson's and Fair's accounts is that for Fair, causation involves transfer of either energy or momentum but not velocity. According to

Salmon's (1994, 1998) Invariant Quantity theory, causation involves an intersection of world lines that results in transmission of an invariant quantity. The proposals of Aronson, Fair, and Salmon come from the philosophy literature. Similar proposals from the psychology literature have been termed generative theories of causation. According to Bullock, Gelman, and Baillargeon (1982), adults believe that causes bring about their effects by a transfer of causal impetus. Shultz (1982) suggests that causation is understood as a transmission between materials or events that results in an effect. According to Michotte (1946/1963), causation is recognized when the parts of an event (e.g., the motions of two objects) constitute a single continuous movement and the motion of the first object extends into the second, what he called an "ampliation of motion." According to Leslie (1994), physical causation is processed by a "Theory of Bodies" (ToBy) that schematizes objects as bearers, transmitters, and recipients of a primitive notion of force.

A recent proposal from the philosophy literature breaks from earlier physicalist models in not requiring a one-way transmission of energy or momentum. According to Dowe's Conserved Quantity Theory (2000), there are two main types of causation: persistence (e.g., inertia causing a spacecraft to move through space) and interactions (e.g., the collision of billiard balls causing each ball to change direction). Causal interactions occur when the world lines (e.g., trajectories) of two objects intersect and there is an *exchange* of conserved quantities (e.g., an exchange of momentum when two billiard balls collide). Unlike transfers, exchanges are not limited to a single direction (e.g., from cause to effect).

Assumptions of physicalist theories. Physicalist approaches to causation share several assumptions. First, they assume that an interaction can be identified as causal on the basis of properties that belong solely to that interaction. Second, defining causal relationships in terms of physical quantities imposes a relatively 'local' level of granularity on the analysis of causal relationships. Transfer of energy, for example, can only occur through local interactions between objects. Third, at the 'local' level of granularity, causal relationships are deterministic (Goldvarg & Johnson-Laird, 2001; Luhmann & Ahn, 2005). The physical quantities that instantiate a direct causal relationships are either present or absent, not present to a probabilistic degree. Fourth, the 'local' nature of causal connections implies that when there is a causal connection between two non-contiguous events, there must be a causal chain of intermediate links, each contiguous to the next (Russell, 1948). Hence, physicalist theories imply the need for causal mechanisms, as has been supported by work in psychology (Ahn & Bailenson, 1996; Ahn & Kalish 2000; Ahn, Kalish, Medin, & Gelman, 1995; see also Bullock et al., 1982; Shultz 1982).

The fifth commonality is that most physicalist theories reduce causal relationships to quantities that cannot be directly observed. In the language of physics, physicalist models hold that people represent causal relationships in terms of their dynamics rather than kinematics. Kinematics concerns the visible properties of an event: the shapes, sizes, positions, points of contact, velocities, and accelerations of the various entities in a particular situation (Schwartz, 1999; Joskowicz & Sacks, 1991; Gilden, 1991). Dynamics, on the other hand, concerns the invisible properties of an event, namely the underlying energies and forces that give rise to the motions.[5] The reduction to invisible quantities reflects the physical priority of invisible quantities over visible quantities. The dynamics of an event are central to people's concept of causation because they are central to causation in the actual world.

Physicalist theories require that people be at least partially aware of dynamic quantities such as force and energy. Such awareness is supported by intuition. When someone picks up a hot pan and then drops it, there is more to the situation than a co-occurrence of events. People feel the energy (the heat) of the pan—not just the pan itself—because the same pan feels different once it cools down. Bigelow, Ellis, and Pargetter (1988) provide a similar example in the case of forces. If something bumps us and we stumble, we feel the force. Again, it is the force that is felt—not only the object—because the same object feels different when it bumps us hard or gently. In physicalist theories—in contrast to dependency theories—energies and forces enter directly into people's representations of causation.[6] However, these representations need not be physically accurate. Clearly, often, they are not (McCloskey, 1983; McCloskey & Kohl, 1983; McCloskey, Washburn, & Felch, 1983). All that is required is a partial sensitivity to their existence.

Importantly, the ability to detect dynamic quantities is possible in principle because of the lawful mapping between kinematics and dynamics, that is, between the visible world of motions and the invisible world of energies and forces. Part of this mapping is captured in Newton's laws of motion. For example, if an object suddenly turns to the right, Newton's 1st law states that the

---

[5] Some physicalist theories ground causation only in kinematics (Michotte & Thines, 1963) or are flexible about whether the representation is either in terms of kinematics or dynamics (Aronson, 1971).

[6] Physicalist theories are compatible with the idea that people's representations of causation might be stored in modality specific areas of the brain, which makes such theories largely compatible with perceptual symbol approaches to representation (Barsalou, 1999; Goldstone & Barsalou, 1998; Robertson & Glenberg, 1998; see also White, 1999).

change in velocity implies acceleration, which entails the presence of a force. Newton's 2nd law, $\mathbf{F} = m\mathbf{a}$, implies that the direction of the force, $\mathbf{F}$, is exactly the same as the direction of acceleration, $\mathbf{a}$. Thus, by observing an instance of change in velocity and the direction of that change, people can, in principle, detect the presence of a force and the direction of its influence.[7] The process of computing forces from kinematics is known as *inverse dynamics*. Work studying people's ability to perform inverse dynamics suggests that they are able to construct at least partial representations of the dynamics of an event (Clement, 1983; Brown & Clement, 1989; Gilden, 1991; Hecht, 1996; Kaiser & Proffitt, 1984; Kaiser, Proffft, Whelan, & Hecht, 1992; Proffitt & Gilden, 1989; Runeson & Frykolm, 1983; Runeson, Juslin, & Olsson, 2000; Runeson & Vedeler, 1993; Twardy & Bingham, 2002).[8]

The sixth assumption of physicalist theories is that physical causation is cognitively more basic than non-physical causation (e.g., social or psychological causation). In support of this assumption, the ability to perceive physical causation begins to develop earlier in infants (around 3 to 4 months) than the ability to perceive social causation (around 6 to 8 months; Leslie, 1994; Cohen, Amsel, Redford, & Casasola, 1998; Oakes, 1994). The final assumption is that non-physical causation is in some way modeled after physical causation (Leslie, 1994; Talmy, 1988). This modeling may occur via a process of analogy in which notions such as "effort" and "intention" are construed of as energies and forces.

Evaluation of physicalist accounts of causations. While the physicalist models discussed so far have advantages over dependency models, they also have several limitations. As with dependency models, current physicalist models are unable to distinguish CAUSE from ENABLE because both CAUSE and ENABLE events are viewed as involving either a transfer or exchange of energy. (See General Discussion for further discussion of this point). Another limitation is that they do not easily represent the concept of PREVENT (Dowe, 2000). If prevention is

---

[7] A particular configuration of forces will produce a particular kinematic pattern, but a particular kinematic pattern need not be associated with only one configuration of forces. This asymmetry between direct and indirect kinematics explains why we can have causal illusions, that is, kinematic patterns that imply forces that are not really there (e.g., Michotte's (1946/1963) launch event).

[8] There has been disagreement over how the process of inverse dynamics might be accomplished in people. According to Runeson and his colleagues, people's perceptual systems allow them to "see" the dynamics of an event via its kinematics, a proposal known as the principle of kinematic specification of dynamics (KSD; see Runeson & Frykolm, 1983; Runeson, Juslin, & Olsson, 2000; Runeson & Vedeler, 1993). Others have suggested that the process of inverse dynamics might be achieved via perceptual heuristics (see Proffitt & Gilden, 1989; Hecht, 1996; Gilden, 1989). The opposite process of computing kinematics from forces is called *direct dynamics*. The psychological correlate of direct dynamics would be the mental simulation of an event from knowledge of forces.

characterized by the lack of transfer or exchange of energy, then it does not differ from the absence of any kind of interaction and if it is characterized by a transfer or exchange of energy, it does not differ from causation. (See Dowe (2000) for an in-depth discussion of the problem of prevention.) The problem with the physicalist models discussed so far is that transmission or exchange of energy is too coarse a criterion for distinguishing causation from other kinds of events that also involve a transmission or exchange of energy. To distinguish causation from other kinds of relationships, a finer level of representation is required.

The dynamics model

The *dynamics model* is a physicalist model of causation. As such, it holds that people represent causal relations in a manner that partially copies or reproduces the way in which causal relationships are instantiated in the real world. It also holds that people can think about non-physical causal relationships by analogy to physical causation. However, unlike other physicalist models, the dynamics model does not associate causation with the transfer or exchange of a physical quantity. Rather, it associates causation with a pattern of forces and a position vector that indicates an endstate. Previous researchers have suggested that causation is closely linked to the notion of force (Ahn & Kalish, 2000, Bigelow et al., 1988; Leslie, 1994). In particular, Bigelow & Pargetter (1990) proposed that causation might be associated with a specific pattern of several forces, though they did not specify the exact pattern. Important parts of the dynamics model are also reflected in diSessa's (1993) phenomenological primitive, Ohm's p-prim,[9] as well as in White's (2000) influence and resistance model, in which causal judgments are likened to the passage of energy in a physical system.

The importance of force in the representation of causation is illustrated by the causal (but static) situations described in (3).[10]

(3) a. Pressure will cause the water to remain liquid at slightly below 0°C.
b. Small ridges cause water to stand on the concrete.
c. The rubber bottom will cause the cup to stay in place.
d. The pole prevented the tent from collapsing.

---

[9] The concept of CAUSE as specified in diSessa's Ohm's p-prim is very similar to that specified in the dynamics model. According to diSessa, Ohm's p-prim is a highly schematized knowledge structure in which an agent that is the locus of an impetus acts against a patient that resists this action but is changed to produce some sort of result. However, unlike the dynamics model, diSessa provides no theoretical machinery for differentiating Ohm's p-prim from, for example, ENABLE, PREVENT or DESPITE (as discussed below), nor did he intend to since his primary concern was in explaining the intuitive sense of mechanism students bring to the task of learning physics.

[10] The sentences in (3) were found through searching Google.

In each of the situations described in (3) nothing happens. Because nothing happens, there is no regular sequence of events, or transfer or exchange of energy, at least at the macro-level. What is present in each of these situations is a configuration of forces. According to the dynamics model, it is this configuration of forces that makes these situations causal (3a-c) or preventative (3d).

The dynamics model is based on Talmy's (1985, 1988) *force dynamics* account of causation (see also Jackendoff, 1991; Pinker, 1989; Siskind, 2000; Verhagen & Kemmer, 1997; Verhagen, 2002; Wolff, 2003; Wolff & Zettergren, 2002). By analyzing the concept of CAUSE into patterns of forces, Talmy showed that the concept of CAUSE could not only be grounded in properties of the world but also be used to define other concepts such as ENABLE, PREVENT, and DESPITE. He also showed how this approach to causation could be extended to many domains of experience, including the physical, intra-psychological, social, and institutional. I incorporate many of Talmy's key ideas into the dynamics model of causation. However, I also introduce several new distinctions and makes significant changes to the theory's underlying semantics. Key differences between the two accounts are summarized in Appendix A.

The dynamics model holds that the concept of CAUSE and related concepts involve interactions between two main entities: an affector and a patient (the entity acted on by the affector). The nature of this interaction can be described at two levels of analysis. The *category level* specifies summary properties of various cause-related concepts. Distinctions at this level are sufficient to distinguish different classes of causal verbs (see Wolff, Klettke, Ventura, & Song, 2005). The *computational level* re-describes the distinctions at the category in terms of units of cognition that represent physical quantities in the world. It is at this level that causes and related concepts are explicitly linked to configurations of force.

The category level of representation. The dynamics model holds that, at the category level, the concept of CAUSE and related concepts can be understood in terms of three dimensions (Wolff & Song, 2003). Specifically, as summarized in Table 1, the concepts of CAUSE, ENABLE, PREVENT, and DESPITE can be captured in terms of 1) the *tendency* of the patient for the endstate, 2) the presence or absence of *concordance* between the affector and the patient, and 3) *progress toward the endstate*. These three binary dimensions allow for eight possible combinations of values. According to the dynamics model, two of these combinations (Y-Y-N & N-Y-Y) do not correspond to a causal relation because they violate the spanning restriction (see below). Two other combinations (N-N-N, N-Y-N) are described by multiple words (see Appendix A). The remaining four possible concepts are listed in Table 1 and discussed below.

Table 1. Representations of CAUSE, ENABLE, PREVENT & DESPITE

	Patient tendency for endstate	Affector-patient concordance	Result: endstate approached
CAUSE	N	N	Y
ENABLE	Y	Y	Y
PREVENT	Y	N	N
DESPITE	Y	N	Y

Note. Y = Yes, N = No

The semantics of these three dimensions are illustrated by the sentences in (4). Consider the example of causation in (4a). In this sentence, the patient (the boat) does not have a tendency for the endstate (heeling). The affector (wind) is not in concordance with the patient and the result occurs. In enabling situations, as in (4b), the tendency of the patient (the body) is for the result (to digest food). The affector (vitamin B) does not oppose the patient, and the result occurs. In preventing situations, as in (4c), the patient (the tar) has a tendency for the result (bonding). The affector (the rain) opposes the tendency of the patient and the result does not occur. In situations where a result occurs despite a certain influence, as in (4d), the tendency of the patient (the river) is towards the result (flooding), and the affector (the dikes) opposes the patient's tendency. In this case, the patient is stronger than the affector and the endstate occurs.

(4) a. Wind caused the boat to heel.
b. Vitamin B enables the body to digest food.
c. Rain prevented the tar from bonding.
d. The river flooded the town despite the dikes.

The category level of representation is supported by recent findings examining the similarity of different causal concepts (Wolff & Song, 2003; Wolff et al., 2002). As indicated in Table 1, the dynamics model predicts that the concepts of CAUSE, ENABLE, and PREVENT are equally similar in meaning since each shares one feature with each other concept: ENABLE and PREVENT both involve patients with a tendency for the result; CAUSE and PREVENT both involve opposition; and CAUSE and ENABLE both lead to results. Therefore, if we were to plot these concepts in a similarity space, they should reside roughly equidistant from each other. In fact, this is exactly what was found when 26 people sorted 48 sentences from the British National Corpus that contained 23 verbs of causation known as periphrastic causative verbs[11] (e.g., the verbs *cause*, *enable,* and *prevent* in (1)) (Wolff & Song, 2003). Participants' sorts were well fit by a two-dimensional MDS solution. As Figure 1 shows, the periphrastic causative verbs in English fall into the three categories predicted by the dynamics model. Importantly, the clusters associated with the three concepts reside roughly equidistant from one another, also as predicted. These results have been replicated with both specific and generic statements of causation. Several rating studies (Wolff et al., 2002; Wolff & Song, 2003), further support the dynamics model's (category level) characterization of causal concepts as differing with respect to the dimensions of tendency, concordance, and result.

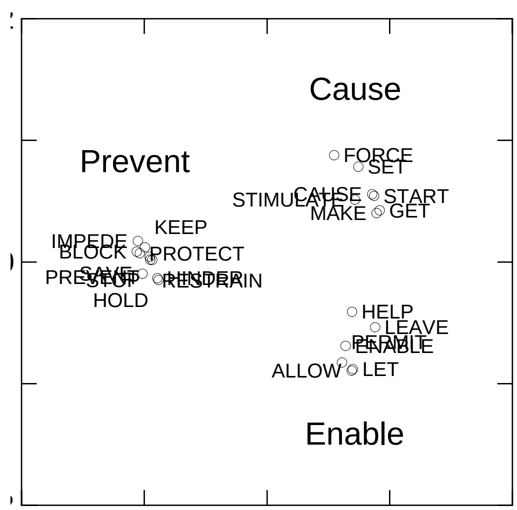


Figure 1. MDS solution of periphrastic causative verbs.

The computational level of analysis. The computational level of the dynamics model re-describes the three dimensions of tendency, concordance, and result in terms of patterns of forces, or vectors. In discussing such vectors I make a distinction between vectors in the world and vectors in people's minds. Vectors in the world are quantities, such as velocity and force, that have a point of origin, a direction, and a magnitude. The vectors in people's representations

[11] Periphrastic causative verbs are sometimes called "pure" causatives since they encode the notion of CAUSE (broadly construed) without specifying a particular result. They behave syntactically and semantically like the verbs *cause*, *enable,* and *prevent* in (1) (Fodor, 1970; Levin & Rappaport Hovav, 1994; Shibatani, 1976; Wolff, 2003; Wolff et al., 2005).

of causation are more qualitative. Specifically, vectors in people's representations are predicted to be relatively accurate with respect to direction, but often imprecise with respect to, though not completely insensitive to, magnitude. People may sometimes be able to infer the relative magnitude of two vectors, that one is greater than another. Uncertainty about the magnitude of the vectors adds a certain amount of indeterminacy to people's representations of force dynamic concepts. It is hypothesized that our mental notion of force vectors includes not only physical forces but also social and psychological forces. Like physical forces, social and psychological forces can be understood as quantities that influence behavior in a certain direction. In this paper, all vectors are in boldface (e.g., **P**). Vectors in the world will be indicated by the subscript "w" (e.g., $\mathbf{P}_w$) while vectors in the mind will be unmarked. Double vertical lines (e.g., $\|\mathbf{P}_w\|$) are used to denote magnitude.

At the computational level, the dynamics model specifies that four types of force vectors are relevant to the mental representation of cause-related concepts. The vector **A** represents the direction of the force that is exerted on the patient by the affector; **P** represents the direction of the force that is generated by the patient itself or, in the absence of such a force, its resistance to change due to frictional or inertial forces; and **O** represents the direction of the force which is based on the summation of the remaining other forces acting on the patient. **R** represents the direction of the force which is the resultant force acting on the patient based on the vector addition of $\mathbf{A}_w$, $\mathbf{P}_w$, and $\mathbf{O}_w$. In addition to these four forces, people's mental representation of the patient's location with respect to an endstate is specified by the vector **E,** which reflects the direction and magnitude of the position vector $\mathbf{E}_w$. When the endstate and patient are points, **E** simply begins at the patient and ends at the endstate, as shown in Figure 2.

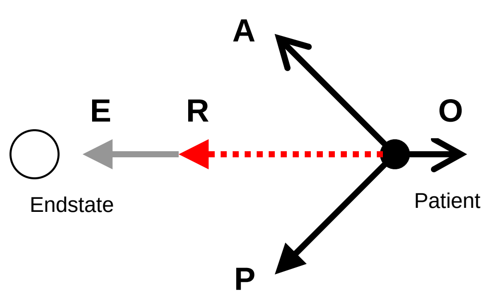


Figure 2. Forces associated with the affector, **A**, patient **P**, and other forces, **O**, combine to produce a resultant force, **R**, that is directed toward the endstate, as specified by the position vector, **E**.

According to the dynamics model, people arrange the vectors **A**, **P**, **O**, **R**, and **E** in a mental structure or schema that resembles a free-body diagram like the one depicted in Figure 1, except that in their mental representations, the magnitudes—indicated by the lengths of the arrows—are relatively uncertain. The circles specify the location of the patient entity with respect to the endstate. The location of the affector entity is not specified because all we need to know is the direction of its influence. It is assumed that the patient is always represented as a point whereas the endstate may be represented as either a point or an area. It is also assumed that changes in state (e.g., melting, breaking, opening) are represented in much the same way as changes of location.[12] When the endstate is an area, the endstate's location can be specified by a set of real 1- or 2-dimensional position vectors, such that every vector from the patient's position to a point that could be considered a part of the endstate would be an element of that set.[13] Finally, it is assumed that in cause-related configurations, ||**A**|| and **||P**|| are greater than 0, but ||**R**|| and ||**O**|| can equal 0.

With these definitions and assumptions in place, the relationship between the category and computational levels of the dynamics model can be specified, as summarized in Table 2.

Table 2. Dimensions in dynamics model

Tendency (of patient for endstate)	**P** & **E** are collinear
Concordance (of affector & patient)	**A** & **P** are collinear
Result: Endstate approached	**R** & **E** are collinear

Tendency - As shown in Table 2, the patient can be viewed as having a tendency for the endstate when the force associated with it, **P,** is in the direction of the endstate, **E**, that is, when **P** and **E** are collinear.[14] For example, in the free-body diagrams illustrating ENABLE, PREVENT and DESPITE in Figure 3, **P** lies in the same direction as **E**, indicating that the patient has a

[12] As proposed in the localist hypothesis, mental and physical states can be viewed as physical locations, and changes in mental and physical states can be construed as motion through space (Anderson, 1971; Lakoff & Johnson, 1980; Langacker, 1986; for a review, see Levin & Rappaport Hovav, 2005). I assume then that restrictions on the representation of motion events likely extend to the representation of changes in state (Pinker, 1989).

[13] In the more general case in which the endstate is other than a point, I expect the definition of concordance must be changed to include a certain level of angular tolerance that would be based, in part, upon the relative size of the target and its proximity to the patient.

[14] What matters in the assessment of collinearity is appearance, not objective reality. Hence, people's assessments of collineararity are not expected to be mathematically perfect.

tendency for the endstate. In the CAUSE configuration, **P** does not point in the same direction as **E**, indicating that the patient does not have a tendency for the endstate.

Concordance - The patient and the affector are in concordance when the patient's tendency, **P**, is in the same direction as the force associated with the affector, **A**, that is, when **P** and **A** are collinear. As shown in Figure 2, collinearity holds in the case of ENABLE but not in the cases of CAUSE, PREVENT, and DESPITE.

Result – The patient will approach the endstate and eventually reach it, barring changes in the forces acting on the patient, when the sum of the forces acting on the patient, **R**, is in the direction of the endstate **E**, that is, when **R** and **E** are collinear. If ||**R**|| equals 0, **R** and **E** are not collinear.

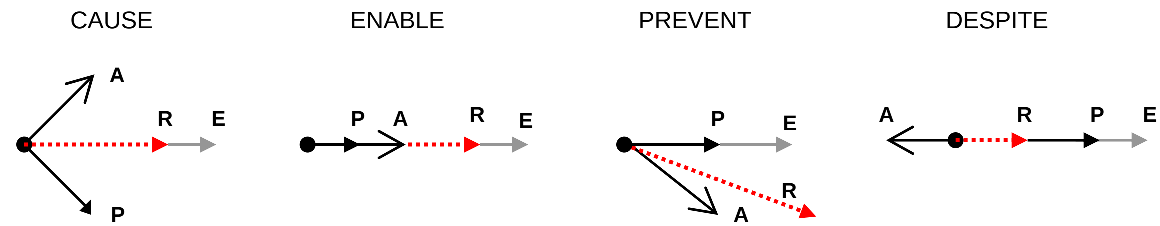


Figure 3. Configurations of forces associated with CAUSE, ENABLE, PREVENT and DESPITE.

Spanning restriction and heuristic. The dynamics model places constraints on what constitutes a valid configuration. Valid configurations are those in which the resultant could be produced from the vector addition of the component vectors. Thus, according to the dynamics model, understanding causal relationships involves evaluating whether **R** reflects the sum of the vectors $\mathbf{A}_{\mathrm{w}}$, $\mathbf{P}_{\mathrm{w}}$, and $\mathbf{O}_{\mathrm{w}}$. People are sensitive to the way in which forces interact in the real world. However, since they represent forces only in terms of **A**, **P**, and **O** (and not $\mathbf{A}_{\mathrm{w}}$, $\mathbf{P}_{\mathrm{w}}$, and $\mathbf{O}_{\mathrm{w}}$), they cannot use exact vector addition to assess **R**. Instead of exact vector addition, I propose that people use a qualitative criterion for deciding whether a resultant could have been produced from the vector addition of two vectors. An implication of the parallelogram law of vector addition is that the resultant of two vectors will always lie on top of or within the region, or *span*,[15] bounded by the vectors being added, as depicted in Figure 4.

[15] The word "span" is used here in a more restricted sense than is used in mathematics. In its usual sense, "span" refers to, for example, the set of resultant vectors, $\mathbf{u}_i$, that can be formed from the equation $\mathbf{u} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2$, where $\mathbf{v}_1$ and $\mathbf{v}_2$ are vectors and c_1 and c_2 are scalars. When using "span" in the context of the dynamics model, I restrict c_1

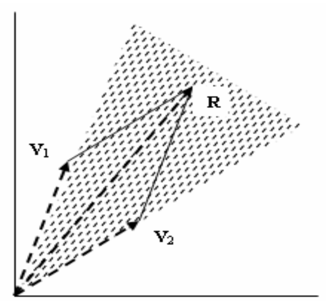


Figure 4. Despite uncertainty about the magnitudes of $\mathbf{V}_1$ and $\mathbf{V}_2$, we can infer that the resultant of the two vectors will reside within the area bounded by $\mathbf{V}_1$ and $\mathbf{V}_2$.

If the resultant lies outside the span of the two vectors being added, the configuration violates the *spanning restriction*. According to the dynamics model, people refer to the spanning restriction in a heuristic—the *spanning heuristic*—to make rough guesses about whether a resultant was produced from the vector addition of the component forces. When a resultant—as indicated by a patient's motion— lies within the span bounded by two vectors, the spanning heuristic warrants the inference that the resultant was produced from the vector addition of the two component vectors. When a resultant lies outside the span, the spanning heuristic holds that the result was not due to the addition of the two component vectors alone.

Testing the dynamics model

Arguably the most important test of a theory of causation is whether it has extensional adequacy. A theory of causation should be able to pick out the range of situations that people judge to be causal while excluding situations that people judge to be non-causal. Unlike dependency models and other physicalist models, the dynamics model makes predictions about what kinds of events will count as causation, as opposed to enablement or prevention. These predictions were tested in the following experiments.

In Experiments 1-4 people viewed animations depicting an affector (e.g., a bank of fans) acting on a patient (e.g., a boat). The motions of the patient were generated by a physics simulator. The inputs to the simulator were the forces associated with the affector and the patient and the patient's mass. In Experiments 5 and 6, the animations depicted intentional forces. A

---

and c_2 to values that are equal to or greater than zero, thus limiting the resultant vectors, $\mathbf{u}_i$, to the region bounded by and including $\mathbf{v}_1$ and $\mathbf{v}_2$.

physics simulator could not be used to implement intentional forces, but because the motions were computer generated, they could be precisely manipulated across conditions.

The animations in Experiments 1-4 showed an inflatable boat, the patient, moving across a shallow pool in relation to a half-submerged cone, the target (see Figure 5). Each animation had two parts. First, the boat moved from the side of the pool to the center, establishing its tendency. Then, a bank of fans (the affector) started blowing. Thus, in the second part of every animation, the force produced by the boat itself combined with the force exerted by the fans to give rise to a resultant force that determined the boat's direction and speed.

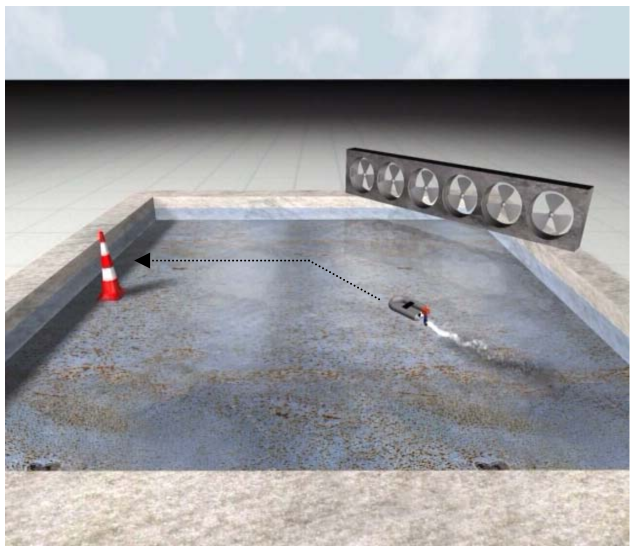


Figure 5: Frame from an animation used to instantiate a CAUSE interaction.

After viewing an animation, participants chose from several linguistic descriptions or "none of the above." All of the descriptions were the same except for the main verb, which was either *caused*, *helped* or *prevented*. I predicted that CAUSE descriptions would be chosen when the boat initially moved away from the cone (Tendency = N), but eventually hit it because of the fans' blowing in the direction of the cone (Concordance = N; Endstate approached = Y). I predicted that ENABLE descriptions would be chosen when the boat moved toward the cone (Tendency = Y), the fans blew in the same direction (Concordance = Y), and the boat ultimately reached the cone (Endstate approached = Y). I predicted that PREVENT descriptions would be chosen when the boat moved toward the cone (Tendency = Y), but the fans blew in another direction (Concordance = N) such that the boat missed the cone (Endstate approached = N).

Finally, I predicted participants would choose the option "none of the above" when none of the above configurations were instantiated.

Experiment 1: One-dimensional interactions

According to the dynamics model, people's causal judgments will be sensitive to the pattern of forces instantiated in a particular animation, and they will be able to make these judgments on the basis of a single observation. This experiment examined the ability of the dynamics model to predict how people would describe configurations of force limited to a single dimension.

Method

Participants. The participants were 18 University of Memphis undergraduates. All participants were native speakers of English.

Materials**.** Eight 3D animations were made from an animation package called Discreet 3ds max 4. The direction and speed of the boat was calculated by a physics simulator called Havok Reactor. In each animation the boat was initially located four boat-lengths away from the center of the pool. In the first half of the animation, the boat moved towards the center, ostensibly under its own power. Once the boat reached the center, the fans began blowing. The animation ended when the boat hit the cone or the side of the pool.

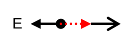
The top of Table 3 shows the direction and magnitudes of the force vectors associated with the affector and patient that were entered into the physics simulator. The affector, $\mathbf{A}_w$, and patient, $\mathbf{P}_w$, vectors were either in the direction of the target or in the opposite direction. In half of the interactions, the affector vector was 1.7 times stronger than the patient vector, while in the remaining interactions the strengths were reversed. Specifically, in configurations 1, 2, 4, and 6, $||\mathbf{A}_w||$ = .984 Newtons and $||\mathbf{P}_w||$ = .59 Newtons. In configurations 3, 5, 7, and 8, $||\mathbf{A}_w||$ = .59N and $||\mathbf{P}_w||$ = .984N. The magnitude of the other forces vector, $\mathbf{O}_w$, was set to 0 Newtons. The boat's mass was 1 kg. The duration of the animations for configurations 1-8 were 17, 8, 6, 17, 10, 5, 10, and 5 seconds respectively.

In the simulated world, the pool was 20 feet by 21 feet. The boat was 1'1" long and 8" wide. The bank of fans was 10' long, 1'9'' high, and 7'' wide. The camera was directed toward the center of the pool at an angle of 25 degrees and was located 10'9'' from the center of the pool. The animations for this and all following experiments can be viewed at http://userwww.service.emory.edu/~pwolff/CLSAnimations.htm.

Procedure**.** The animations were presented in random order on Windows-based computers. After each animation, participants chose a sentence that best described the occurrence. All of the sentences were the same (*The fans ____ the boat to [from] hit[ting] the cone*) except for the verb, which was either *caused*, *helped* or *prevented*. Another option was *none of the above*. Participants indicated their answers by clicking a radio button next to their choice.

Design. Participants saw all eight animations. There were two factors: ConfigType (CAUSE, ENABLE, PREVENT, UNSPECIFIED) and ResponseType (*Cause*, *Help*, *Prevent*, *No verb*).

Table 3. Experiment 1 predictions and results by configuration and response type (mean (SD))

Config. #	1 N-N-Y	2 Y-Y-Y	3 Y-Y-Y	4 Y-N-N	5 Y-N-Y	6 N-Y-N	7 N-N-N	8 N-Y-N
Affector ( ) Patient ( ) Result. ( )								
Predicted	CAUSE	ENABLE	ENABLE	PREVENT	Unspecified	Unspecified	Unspecified	Unspecified
"Cause"	**.94** (.236)	.11 (.323)	.06 (.236)	-	-	-	-	-
"Help"	.06 (.240)	**.89** (.323)	**.94** (.236)	-	.11 (.323)	-	.06 (.236)	-
"Prevent"	-	-	-	**1** (0)	.06 (.236)	-	-	.06 (.236)
"No verb"	-	-	-	-	**.83** (.383)	**1** (0)	**.94** (.236)	**.94** (.236)

Results and discussion

The key question addressed in this experiment was whether the dynamics model could predict which situations would be judged to be causal, enabling, preventing or unclassifiable. The predictions of the dynamics model were fully borne out by the results. The bottom of Table 3 shows the percentage of times people chose each of the four options for each configuration of forces The results were analyzed using log-linear modeling. Like ANOVA, log-linear modeling can be used to test for main effects and interactions between those main effects. A log-linear model based on the factors ConfigType (4), and ResponseType (4) and their two-way interaction was fitted to the observed frequencies.[16] A Pearson's chi-square implied that such a model agreed well with the observed frequencies, as there was no evidence for a difference between the predictions of the model and the observed frequencies, $\chi^2(3, N = 144) = .158, p < .984$.

Each factor and interaction was removed from this model to examine its relative contribution to the model's fit. As predicted, removing the interaction between ConfigType and ResponseType from the model resulted in a significant decrease in the fit, $\chi^2(9, N = 144) =$

[16] In all experiments, cells with a frequency of zero were randomly assigned the values 1 or 2.

111.87, $p < .0005$. This interaction indicates that people provided different responses for different configurations. Removal of the main factors of ConfigurationType, $\chi^2(3, N = 144) = 1.56$, $p = .669$ and ResponseType, $\chi^2(3, N = 144) = .16$, $p = .983$, did not have a significant effect on the fit of the model.

As predicted, people chose the sentence containing *cause* to describe the CAUSE configuration, $\chi^2(3, N = 18) = 35.19$, $p < .0005$; the sentence containing *help* to describe the ENABLE configuration, $\chi^2(3, N = 18) = 58.70$, $p < .0005$; the sentence containing *prevent* to describe the PREVENT configuration, $\chi^2(3, N = 18) = 32$, $p < .0005$; and the option "none of the above" for the unspecified configurations, $\chi^2(3, N = 72) = 153.60$, $p < .0005$. The results demonstrate that the dynamics model is able to differentiate related causal concepts and that causal relations can be apprehended from a single observation.

Importantly, the results indicate that people's categorizations of force configurations are based on two forces, not just one. For example, if people's judgments were based only on the affector force, they would have chosen *prevent* whenever the boat missed the cone (6-8). Instead, *prevent* was restricted to cases in which the boat had an initial tendency for the endstate (4)—as indicated by the patient's force—just as predicted by the model. Likewise, people would have chosen either *cause* or *help* when the boat hit the cone (5), but they did not. In configuration 5, *help* was not chosen, according to the model, because the affector and patient vectors were in opposition; *cause* was not chosen, according to the model, because the patient had a tendency for the endstate. People's causal judgments clearly involved multiple forces.

The results show that the dynamics model can explain the difference between causation and other concepts for interactions occurring within a single dimension. Interestingly, the model easily extends to situations across two dimensions, as is examined in the next experiment.

Experiment 2: Two-dimensional interactions

The procedures in Experiment 2 were the same as in Experiment 1. The materials were the same as well, except that the angles between the forces associated with the affector and patient varied from 0° to 180° degrees in 45° increments. It was predicted that peoples' descriptions would match the force configurations hypothesized to instantiate the different causal relations—or, in the case of the unspecified configurations, people would choose "none of the above."

Method

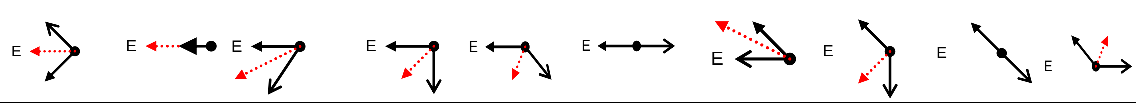
Participants. The participants were 18 University of Memphis undergraduates. All participants were native speakers of English.

Materials. Ten 3D animations were the same as in Experiment 1 except that the affector and patient force vectors were oriented in several directions other than directly towards or away from the target, and the magnitudes of $\mathbf{A}_w$ and $\mathbf{P}_w$ were always the same (.59 Newtons). The ten vector configurations at the top of Table 4 include five in which the patient vector is oriented away from the target by 45° (1, 7-10) and five in which the patient vector is oriented towards the target (2-6). The affector vector was rotated around the patient from 180° (i.e., in the direction opposite to the endstate) to 360° (in the same direction as the endstate) in 45° increments. The size of the various elements in the scene and the location of the simulated camera were the same as in Experiment 1. Animations 1-4 and 6-7 were 5 seconds long; animations 5 and 9 lasted 6 seconds; animation 10 lasted 7 seconds; and animation 8 lasted 8 seconds.

Procedure The procedure was the same as in Experiment 1.

Design. Participants saw all ten animations. There were two factors: ConfigType (CAUSE, ENABLE, PREVENT, UNSPECIFIED) and ResponseType (*Cause*, *Help*, *Prevent*, *No verb*).

Table 4.
Experiment 2 predictions and results by configuration and response type (mean (SD))

Config. #	1	2	3	4	5	6	7	8	9	10
	N-N-Y	Y-Y-Y	Y-N-N	Y-N-N	Y-N-N	Y-N-N	N-N-N	N-N-N	N-N-N	N-N-N
Affector ( ) Patient ( ) Result. ( )										
Predicted	CAUSE	ENABLE	PREVNT	PREVNT	PREVNT	PREVNT	Unspecified	Unspecified	Unspecified	Unspecified
"Cause"	**.89** (.323)	.11 (.323)	-	-	-	-	-	-	-	-
"Help"	.11 (.323)	**.83** (.384)	-	-	-	-	-	-	-	-
"Prevent"	-	-	**.94** (.236)	**.94** (.236)	**.89** (.323)	**.89** (.323)	-	.17 (.383)	-	.11 (.323)
"No verb"	-	.06 (.236)	.06 (.236)	.06 (.236)	.11 (.323)	.11 (.323)	**1** (0)	**.83** (.384)	**1** (0)	**.89** (.323)

Results and discussion

The key question addressed in this experiment is whether the dynamics model can account for peoples' judgments of causation when the forces span two dimensions. The predictions of the dynamics model were supported once again. The lower portion of Table 4 shows the percentage of times people chose each of the four possible options for each configuration of force. The results provide further evidence that CAUSE and related concepts are determined on the basis of multiple forces. If causal judgments were based on only one force, there would be no basis for distinguishing CAUSE from ENABLE, nor any basis for distinguishing the PREVENT from the non-classifiable situations in which the boat did not hit the cone (configurations 7-10). Not only do the results demonstrate that the dynamics model can distinguish CAUSE from other relations, they also show that the model can distinguish causation from non-causation.

The above conclusions are supported by log-linear modeling. A log-linear model based on the factors ConfigType (4), and ResponseType (4) and their two-way interaction was fitted to the observed frequencies. A Pearson's chi-square indicated that such a model agreed well with the observed frequencies, as implied by the lack of a difference between the predictions of the full model and the observed frequencies, $\chi^2(3, N = 180) = .233$, $p < .972$.

Each factor and interaction was removed from this model to examine their relative contribution to the model's fit. As predicted, removing the interaction between ConfigType and ResponseType from the model resulted in a significant decrease in the fit, $\chi^2(9, N = 180) = 105.71$, $p < .0005$. This interaction indicates that people responded differently to different configurations. Removal of the main factors of ConfigType, $\chi^2(3, N = 180) = .23$, $p = .972$, and ResponseType, $\chi^2(3, N = 180) = .16$, $p = .984$, did not have a significant effect on the fit of the model.

As predicted, people chose the sentence containing *cause* to describe the CAUSE configuration (1), $\chi^2(3, N = 18) = 29.48$, $p < .0005$; the sentence containing *help* to describe the ENABLE configuration (2), $\chi^2(3, N = 18) = 29.63$, $p < .0005$; the sentence containing *prevent* to describe the PREVENT configurations (3-6), $\chi^2(3, N = 72) = 35.19$, $p < .0005$; and the option "none of the above" for the unspecified configurations (7-10), $\chi^2(3, N = 72) = 32.182$, $p < .0005$.

While the results so far support the dynamics model, several concerns could be raised. First, in Experiments 1 and 2, only three verbs were considered: *cause*, *help*, and *prevent*. If the dynamics model is an account of how people represent the concepts of CAUSE, ENABLE, and PREVENT in general and not just the meanings of three particular verbs, similar results should obtain for other verbs of causation such as *get*, *make*, *enable*, *block* or *keep*. Likewise, the dynamics model should be shown to extend to scenarios other than the one used in Experiments 1 and 2. Finally, as previously noted, the dynamics model predicts another concept that has not yet been considered, namely, DESPITE. As listed in Table 1, the concept of DESPITE is associated with a patient that has a tendency for the endstate, an affector is not concordant with the patient, and a result that occurs (see Figure 3). Such configurations seem to be implied in such sentences as *The boat reached the shore despite the current* or *The bird overcame the high winds to reach her nest.*

The following experiment addressed these concerns. First, the verbs used in Experiments 1 and 2 were each replaced by three different verbs. The verbs were chosen from among the CAUSE, ENABLE, and PREVENT verbs studied in Wolff and Song (2003). Participants viewed and chose descriptions for four different scenarios. Finally, the description choices included a DESPITE option.

I predicted that for the CAUSE configurations, people would choose descriptions based on different CAUSE verbs; for the ENABLE configurations, people would choose descriptions based on ENABLE verbs; for the PREVENT configurations, people would choose descriptions based on PREVENT verbs; and for the DESPITE configurations, people would choose descriptions containing *despite*. For the unspecified configurations, I predicted that participants would choose the *none-of-the-above* option. This pattern of results was predicted to occur across the four scenarios.

I also predicted that the pattern of results would be the same for the different CAUSE and PREVENT verbs. However, for the ENABLE verbs, I predicted that for this set of animations, people would prefer to use the verb *help* to the verbs *enable* and *let*. As suggested by the results in Experiments 1 and 2, the verb *help* can be used to describe a situation in which the affector assists the patient towards an endstate even though the patient may be able to reach the endstate without this assistance. For example, helping someone finish his homework does not necessarily imply that he cannot finish it on his own. In contrast, *enable* and *let* seem to imply that the patient cannot reach the endstate without the aid of the affector (Goldvarg & Johnson-Laird, 2001). This additional implication was not warranted in the case of the animations used in Experiments 1 and 2 or in the animations in the next experiment. I expected, then, that people would be more willing to use *help* than *enable* or *let* for the animations in the next experiment.

Experiment 3

The methods and materials were similar to those used in Experiments 1 and 2, except that in this experiment, four different scenarios were presented, several different CAUSE, ENABLE, and PREVENT verbs were used, and a DESPITE option was added to the list of descriptions.

Method

Participants. The participants were 27 University of Memphis undergraduates. All participants were native speakers of English.

Materials. The experiment involved twenty-four animations that were based on one of four scenarios: 1) a blimp moving with respect to a docking tower, 2) an ice boat moving with respect to a bonfire, 3) a helicopter moving with respect to a landing pad, and 4) a boat moving with respect to a cone in a pool. Sample frames of the four scenarios are depicted in Figure 6.

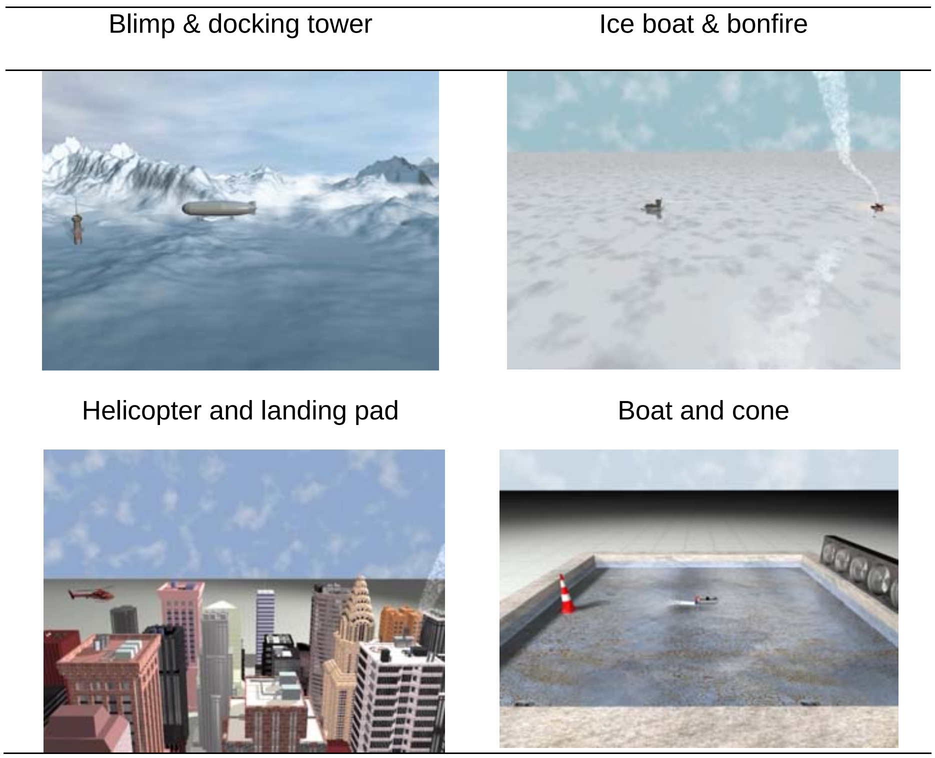


Figure 6: Sample frames from the different kinds of animations used in Experiment 3.

In each animation, the patient was initially located four patient-lengths away from the center of the scene. In the first half of the animation, the patient moved towards the center under its own power. Once the patient reached the center, the wind or the fans started blowing. Wind was indicated by moving smoke or water vapor. The animation ended when the patient reached the endstate or neared the side of the screen.

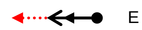
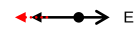
The directions of the forces entered into the physics simulator are shown in Table 5 and the magnitude of these forces are shown in Appendix B. The blimp animations were 11, 6, 6, 9, 4, and 8 seconds for configurations 1-6, respectively. The ice boat animations were 8, 3, 6, 7, 3, and

6 seconds long for configurations 1-6, respectively. The helicopter animations were 10, 4, 6, 7, 4, and 6 seconds for configurations 1-6, respectively. Finally, the motor boat animations were 16, 6, 6, 10, 6, and 9 seconds for configurations 1-6, respectively.

Procedure**.** The procedure was similar to that used in Experiments 1 and 2. The animations were presented in random order on Windows-based computers. After each animation, participants chose a sentence that best described the occurrence. All of the sentences were structurally the same (*The wind* **____** *the blimp to [from] reach[ing] the docking tower*) except for the verb. In the CAUSE sentences, the verbs used were *cause*, *make*, or *get*. ENABLE sentences included *enable*, *help*, or *let*, and PREVENT sentences included *prevent*, *block* or *keep*. The DESPITE option always used the preposition *despite* (e.g., *The blimp reached the docking tower despite the wind*). The order of the sentence choices varied randomly for each animation except for the last choice, which was always *none of the above*. Participants indicated their answers by clicking a radio button next to their choice.

Design. Participants saw all twenty-four animations, which instantiated the six configurations displayed in Table 5 for each of the four scenarios displayed in Figure 6. There are 27 combinations of verbs (3^3) that can be formed from three CAUSE, ENABLE, and PREVENT verbs. Each participant saw sentences based on one of these combinations for all of the animations he or she saw. There were four factors: ConfigType (CAUSE, ENABLE, PREVENT, DESPITE, UNSPECIFIED), Scenario (Blimp, Ice boat, Helicopter, Raft), ResponseType (Cause, Enable, Prevent, Despite, NoVerb), and VerbType (e.g., *cause*, *enable*, *prevent*).

Table 5. Experiment 3 predictions and results by configuration and response type (mean (SD))

Configuration # Type	1 CAUSE N-N-Y	2 ENABLE Y-Y-Y	3 PREVENT Y-N-N	4 DESPITE Y-N-Y	5 Unspecified N-Y-N	6 Unspecified N-N-N
Affector ( ) Patient ( ) Result. ( )						
CAUSE	**.84** (.366)	.19 (.398)	-	.02 (.135)	-	-
Cause	.28 (.450)	.06 (.247)	-	.02 (.135)	-	-
Make	.29 (.459)	.06 (.230)	-	-	-	-
Get	.27 (.445)	.07 (.263)	-	-	-	-
ENABLE	.03 (.165)	**.69** (.467)	.04 (.211)	.01 (.096)	.03 (.165)	-
Enable	.01 (.096)	.19 (.390)	.04 (.211)	-	.01 (.096)	-
Help	.02 (.135)	.37 (.485)	-	-	.01 (.096)	-
Let	-	.13 (.337)	-	.01 (.096)	.01 (.096)	-
PREVENT	.05 (.211)	-	**.91** (.291)	-	.11 (.316)	.05 (.211)
Prevent	.02 (.135)	-	.29 (.454)	-	.04 (.190)	.03 (.165)
Keep	.01 (.096)	-	.30 (.459)	-	.04 (.190)	.02 (.135)
Block	.02 (.135)	-	.32 (.470)	-	.04 (.190)	-
DESPITE						
Despite	.01 (.096)	.04 (.190)	.02 (.135)	**.96** (.190)	-	.04 (.190)
None of the above	.07 (.263)	.08 (.278)	.03 (.165)	.01 (.096)	**.86** (.347)	**.91** (.278)

Results

The predictions of the dynamics model were supported once again. Table 5 shows the proportion of times people chose each of the five possible response types (CAUSE, ENABLE, PREVENT, DESPITE, NOVERB) broken down by verb for each of the vector configurations.[17] People chose CAUSE verbs for the CAUSE configuration, ENABLE verbs for the ENABLE configuration, PREVENT verbs for the PREVENT configuration, the preposition *despite* for the DESPITE configuration, and *none of the above* for the remaining configurations.

The above conclusions are supported by log-linear modeling. A log-linear model based on the factors ConfigType (5), Scenario (4), ResponseType (5) and their two- and three-way interactions was fitted to the observed frequencies. A Pearson's chi-square indicated that such a model agreed well with the observed frequencies, as implied by the lack of a difference between the predicted frequencies of the model and the observed frequencies, $\chi^2(48, N = 540) = 16.91$, *p*

[17] In the following analyses, responses to the two animations with unspecified configurations are averaged together to make the frequencies of responses to these configurations comparable with the other types of configurations that are instantiated by only one animation.

$> .999$. Each factor and interaction was removed from this model to examine its relative contribution to the model's fit. As predicted, removing the interaction between ConfigType and ResponseType from the model resulted in a significant decrease in the fit, $\chi^2(16, N = 540) = 886.26$, $p < .0005$. This interaction indicates that responses differed across different configuration types. Removal of the remaining two-way interactions (ScenarioType*ResponseType, $\chi^2(12, N = 540) = 5.61$, $p = .934$, ScenarioType* ConfigType, $\chi^2(12, N = 540) = 3.31$, $p = .993$, and the one three-way interaction, ConfigType *ResponseType*ScenarioType, $\chi^2(48, N = 540) = 16.45$, $p > .999$, did not have a significant effect on the fit of the model. Removal of the main factors of ConfigType, $\chi^2(4, N = 540) = 1.68$, $p = .794$, ResponseType, $\chi^2(4, N = 540) = 2.37$, $p = .668$, did not have a significant effect on the fit of the model either. Importantly, removing the main factor of ScenarioType, $\chi^2(4, N = 540) = 0.47$, $p = .926$, had no appreciable effect on the fit of the model. The nonsignificance of this factor and its associated interactions suggests that people treated the various scenarios as essentially the same, as predicted. The results indicate that the most important factor in people's choice of description was not the specific content of the animations, but rather the underlying configuration of force.

Combining the responses across scenarios, the Pearson's chi-square indicated that participants chose CAUSE descriptions for the CAUSE configuration (1), $\chi^2(4, N = 108) = 279.96$, $p < .0005$; ENABLE descriptions for the ENABLE configuration (2), $\chi^2(4, N = 108) = 127.13$, $p < .0005$; PREVENT descriptions for the PREVENT configuration, (3), $\chi^2(4, N = 108) = 333.34$, $p < .0005$; and DESPITE descriptions for the DESPITE configurations (4, N = 108), $\chi^2(4, N = 108) = 387.47$, $p < .0005$. Finally, participants chose *none of the above* for the unspecified configurations (5), $\chi^2(4, N = 108) = 290.18$, $p < .0005$, and 6, $\chi^2(4, N = 108) = 337.46$, $p < .0005$ (see Table 5).

It was expected that the dynamics model would be able to capture the conceptual commonalities shared by the verbs that encode the general notions of CAUSE, ENABLE, and PREVENT. In support of this, participants showed no preference for one of the CAUSE verbs over the others, $\chi^2(2, N = 108) = 0.154$, $p = .926$. In addition, in describing the PREVENT configurations, participants showed no preference for one of the PREVENT verbs over the others, $\chi^2(2, N = 108) = 0.265$, $p = .876$. However, in the case of the ENABLE configurations, participants showed a stronger preference for the verb *help* than for the verbs *enable* and *let*, $\chi^2(2, N = 108) = 15.027$, $p = .001$. In the animations used in this experiment, the patient appeared likely to reach the goal even without the presence of the affector, which, as discussed earlier, reflects the special semantics of the verb *help*, whereby the result could be achieved by the patient itself. In contrast, the verbs *enable* and *let* imply situations in which the outcome is unlikely to occur unless the affector is present. Nevertheless, all three verbs, as well as other ENABLE verbs imply a patient with a tendency for the endstate and an affector that does not oppose it. Because of these common elements of meaning, people were often willing to use *enable* and *let* to describe the ENABLE animations though not as frequently as *help* (see Table 5). Differences in the meanings of ENABLE verbs will be discussed further in reference to Experiment 6.

The results from Experiments 1-3 show that the dynamics model, unlike dependency and other physicalist models meets the criterion of extensional adequacy. The results also support the dynamics model's account of how people determine causation on the basis of a single observation. According to the model, people identify causal relationships by constructing representations of the forces acting on the patient. However, the data so far are open to an alternative possibility; specifically, they could be explained in terms of kinematics rather than dynamics. In a kinematics account, only visible movements—specifically, the velocities—are considered in the classification of interactions. For example, causation might be defined as an interaction in which the patient was not moving toward the endstate, but then moved toward the endstate once the affector made contact with it. Enablement could be defined as an interaction in which the patient was moving toward the endstate, but then moved more quickly toward the endstate once the affector made contact with it. Finally, prevention might be defined as an interaction in which the patient was moving toward the endstate, but then moved away from the endstate once the affector made contact with it.

One way to test between kinematics and dynamic approaches to causation would be to examine whether people are aware of the way in which forces are added. If people's causal judgments are based on the dynamics of an event, they should be relatively sensitive to motions that do not conform to the way forces are added. On the other hand, if people's causal judgments are based on kinematics, peoples' causal judgments should be insensitive to such violations.

As discussed earlier, it is assumed that people use a qualitative criterion, *the spanning heuristic*, to determine whether a particular resultant could be derived from a particular set of forces. Thus, when a patient moves in a direction that lies within the area between the forces acting on the patient (see Figure 4), the spanning heuristic should lead people to assume that the resultant is produced from the vector addition of those forces. Conversely, when the resultant does not reside within the span of the component vectors, it can be said that the configuration violates the *spanning restriction*.

The spanning heuristic provides a rough method of evaluating whether the net force acting on a patient is derivable from the overt forces acting on the patient. However, in certain circumstances, the heuristic may lead people to incorrectly infer that the net force acting on the patient is fully explained in terms of the perceived forces when, in fact, there are other forces in play. Such an illusion of sufficiency is most likely to occur when there is more than one external force acting on the patient, that is, when $||\mathbf{O_w}|| > 0N$. For example, consider the three scenes and free-body diagrams in Figure 7. The forces entered into the physics simulator for all the scenes are depicted in the first free-body diagram. In the first animation, a boat motors to the middle of a pool, two sets of fans turn on, and the boat moves toward the cone and ultimately hits it. The second panel shows a frame from an animation that is exactly the same as the one on the left except that the one of the fans is not shown (though its force is still present). In this animation, the boat moves into the area bounded by the overt forces; hence, according to the spanning heuristic, the fan may be construed as a cause of the boat's hitting the cone.

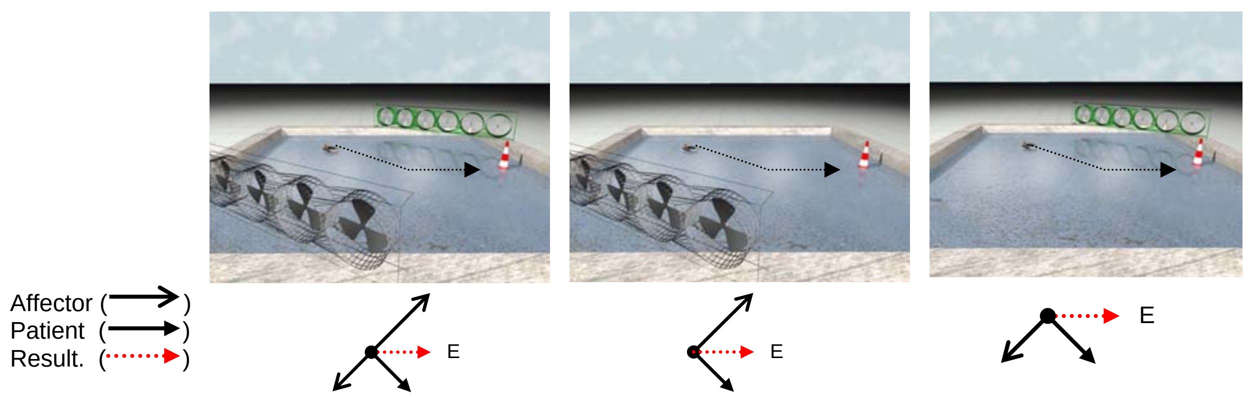


Figure 7. In each animation, the boat motors to the middle, the fans turn on, the boat changes course, and the boat hits the cone. Each animation is based on the same configuration of forces

as shown in the first panel. However, in the second and third panels, only one of the two fans appears in the animation, as implied by the incomplete.

The third panel shows an animation that is also exactly the same as the one in the first panel except that the opposite fan is not shown. In this scene, based on single visible fan, the boat's direction lies outside the area bounded by the perceived forces. According to the spanning heuristic, then, the visible fan cannot be construed as a cause of the boat's hitting the cone. The idea that people's causal judgments are sensitive to the resultant forces acting on the patient was tested in the next experiment.

Experiment 4

This experiment examined whether people's causal judgments are based on dynamics or kinematics. I predicted that if peoples' representations of causation are based on dynamics, people should be sensitive to violations in the way the forces are added in a situation; otherwise, they should not be sensitive to such violations.

Participants saw four pairs of animations like those in the second and third panels in Figure 7 (both instantiating CAUSE configurations). In all of the animations, two affector forces were in play, but only one of the forces was shown. In each pair of animations, one animation depicted a situation in which the resultant was within the span of the overt forces, and the other depicted a situation in which the resultant was not within the span. Participants were to indicate whether the fan in the animation "caused" the boat to hit the cone. Per the spanning heuristic, it was predicted that when the resultant lay within the span of the forces associated with the boat and the fans, people would agree that the fans caused the boat to hit the cone. It was also predicted that when the resultant was outside the span of the overt forces, people would indicate that the fans did not cause the boat to hit the cone.

Method

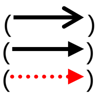
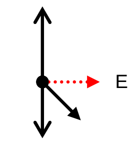
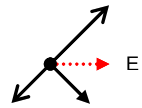
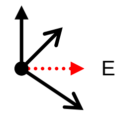
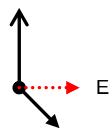
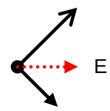
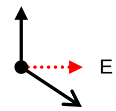
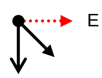
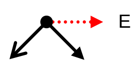
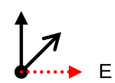
Participants. The participants were 20 Emory University undergraduates. All participants were native speakers of English.

Materials. Four pairs of 3D animations depicting boats and fans were constructed from the four base configurations shown in the top row of Table 6. In each animation, three forces acted on the patient: the force generated by the patient itself, $\mathbf{P}_{\mathbf{w}}$, and two external forces, $\mathbf{A}_{\mathbf{w}1}$ and $\mathbf{A}_{\mathbf{w}2}$, which were ostensibly generated by the banks of fans (see Figure 7). The two animations in each pair were constructed from the same set of forces and showed only one bank of fans; the animations differed in which of the two banks was shown. In one animation, the spanning restriction was honored: the resultant—as indicated by the direction of the boat after the fans started blowing—was within the area bounded by the force associated with the visible bank of fans and the force associated with the boat. In the other member of each pair, the spanning restriction was not honored: the resultant LAY outside the area bounded by the forces associated with the visible bank of fans and the boat.

The directions of the forces entered into the physics simulator are given in Table 6. More precisely, in configuration 1, the angles between $\mathbf{A}_{w1}$ and $\mathbf{E}_w$, $\mathbf{P}_w$ and $\mathbf{E}_w$, and $\mathbf{A}_{w2}$ and $\mathbf{E}_w$ were 90°, 45°, and 90°, respectively. In configuration 2, the angles between $\mathbf{A}_{w1}$ and $\mathbf{E}_w$, $\mathbf{P}_w$ and $\mathbf{E}_w$, and $\mathbf{A}_{w2}$ and $\mathbf{E}_w$ were 45°, 45°, and 135°, respectively. In configuration 3, the angles between $\mathbf{A}_{w1}$ and $\mathbf{E}_w$, $\mathbf{P}_w$ and $\mathbf{E}_w$, and $\mathbf{A}_{w2}$ and $\mathbf{E}_w$ were 180°, 45°, and 45°, respectively. In configuration 4, the angles between $\mathbf{A}_{w1}$ and $\mathbf{E}_w$, $\mathbf{P}_w$ and $\mathbf{E}_w$, and $\mathbf{A}_{w2}$ and $\mathbf{E}_w$ were 90°, 45°, and 45°, respectively. The magnitudes of the forces are shown in Table 6. The duration of the animations based on configurations 1-4 were 7, 6, 20, and 8 seconds respectively.

As in previous animations, the BOAT was initially located four BOAT-lengths away from the center of the pool. The patient moved towards the center of the pool under its own power. Once the patient reached the center, the fans started blowing. The animation ended when the patient hit the endstate or neared the side of the screen. To help people see the directions of the boats, a second smaller animation was included in the upper right-hand corner of the main animation. In this second animation, the same scene was shown from a simulated camera angle that was directly above the pool and looking down.

Table 6. Experiment 4 predictions and results by configuration and response type (mean (SD))

Configuration #	1	2	3	4
	Base configurations			
Affector ( ) Patient ( ) Result. ( ) 				
Magnitudes	$\|\|\mathbf{A}_{w1}\|\| = 1.4N$ $\|\|\mathbf{A}_{w2}\|\| = 1N$ $\|\|\mathbf{P}_{w}\|\| = 1N$	$\|\|\mathbf{A}_{w1}\|\| = 2N$, $\|\|\mathbf{A}_{w2}\|\| = 1N$, $\|\|\mathbf{P}_{w}\|\| = 1N$	$\|\|\mathbf{A}_{w1}\|\| = 1N$ $\|\|\mathbf{A}_{w2}\|\| = 1N$ $\|\|\mathbf{P}_{w}\|\| = 1N$	$\|\|\mathbf{A}_{w1}\|\| = 1N$ $\|\|\mathbf{A}_{w2}\|\| = 2.4N$ $\|\|\mathbf{P}_{w}\|\| = 1N$
	Within span variant (affector = $\mathbf{A}_1$)			
				
Proportion "Yes" cause	.75 (.444)	.85 (.366)	1 (0)	.75 (.444)
	Outside of span variant (affector = $\mathbf{A}_2$)			
				
Proportion "Yes" cause	.2 (.410)	.15 (.366)	0 (0)	.35 (.489)

Procedure. Participants saw all eight animations in random order on Windows-based computers. After each animation, participants responded "*Yes*" or "*No*" to the question "*Did the fans cause the boat to hit the cone?*" Participants were then asked to *"[r]ate how confident you are in your answer*" on a five-point Likert scale (not confident, slightly confident, confident, very confident, extremely confident). Participants indicated their answers by clicking a radio button next to their choice. Participants progressed through the animations at their own pace and could repeat an animation as many times as they wanted.

Design. The main factor of SpanType (spanning honored; spanning not honored) was run within participants.

Results and Discussion

The results indicated that people were sensitive to violations in the adding of the overt forces. When the boat moved in a direction that was consistent with the spanning restriction, participants were quite willing to say that the fan "caused" the boat to hit the cone (M = .838, SD = .203).

However, when the boat moved in a direction that was not consistent with the spanning restriction, participants were not willing to say that the fan “caused” the boat to hit the cone (M = .175, SD = .282).

This observation was supported by *t*-tests that showed that the proportion of responses affirming that the fans caused the boat to hit the cone was higher in the spanning condition than in the no-spanning condition across both participants, $t(19) = 9.668$, $p < .0005$, and items, $t(6) = 7.10$, $p < .0005$. Importantly, the difference between the spanning and no-spanning conditions was not due to differences in uncertainty about how to classify the events. Participants in both the spanning (M = 3.73, SD = .884) and no-spanning conditions (M = 4.05, SD = .746) indicated that they were “very confident” in their causal judgments. Participants’ levels of confidence were marginally higher in the no-spanning condition than in the spanning condition across participants, $t(19)=2.08$, $p = .051$, but not across items, $t(6) = 1.21$, $p = .273$.

The results are not easily explained if people based their causal judgments only on the kinematics of the scenes. In terms of kinematics, causation might be defined as occurring when the patient was not moving towards the endstate, but then moved towards the endstate once the affector made contact with it. According to this definition, the fans should have been considered a cause in both the spanning and non-spanning conditions, but they were chosen as causes only in the spanning condition. It could be noted that the patient moved away from the fans in all of the spanning conditions, and that in three of the four non-spanning conditions, the boat moved towards the fans. However, in the fourth pair of animations, the boat moved away from the fans in both conditions. Even when the boat moved away from the fans, people viewed the fans as non-causal, presumably because the boat’s motion was not consistent with the way the forces could be added.

In addition to these results, several other problems remain for a kinematics account of causation. As discussed above, the concept of CAUSE extends to situations in which there are conflicting forces, but no change occurs (e.g., *Pressure can cause water to remain liquid at slightly below 0°C*). A kinematics-based account cannot motivate why these situations can be viewed as causal nor can it distinguish such situations from static, non-causal situations (e.g., **The tree causes the roof to be under the branch*). Another limitation to a kinematics approach is that it does not easily explain our language for non-physical causation. In describing social causation, we rarely talk about "social velocities" or "peer accelerations." Rather, we talk about "social forces" and "peer pressures."

The results also support the dynamics model's assumption that people's judgments of causation do not require knowing the exact magnitudes of the forces. In the spanning conditions, the animations did not provide enough information to determine whether the boat's course was due to the force associated with just the one fan or due to that force in combination with (an)other hidden force(s). Regardless, when the boat moved within the span of the overt forces, people agreed that the fan "caused" the boat to hit the cone. Knowledge of the precise magnitudes is not necessary for classifying the situation as causal. What appears to be necessary is awareness of the direction of the forces, which supports the hypothesis that people think about causation in terms of vectors.

Causation involving physical and non-physical forces. The results so far suggest that the dynamics model meets the criterion of extensional adequacy at least for purely physical events. However, non-physical forces in causation are extremely common. If the dynamics model is a general model of causation, we need to know whether its ability to explain causal judgments extends to causal relationships that involve non-physical influences, including intentions, desires, and social directives. As discussed previously, such influences can be construed of as forces since they can be viewed as having an origin, direction and magnitude (Copley, in press). Talmy (1988) also has argued that intentions and desires could be treated as roughly analogous to physical forces, as illustrated in sentences like the ones in (5).

(5) a. Ice caused the branches to bend. (physical forces only)

b. Seeing the ice caused Michelle to stay home. (psychological forces)

c. Michelle caused Tom to stay home by telling him about the ice. (social forces)

d. Ice storm warnings caused schools in Atlanta to close. (institutional forces)

The causation implied in (5a) involves physical forces only. The causation in (5b) implies psychological forces only. Michelle's tendency to go to work is opposed by the realization that it would be unwise to travel with ice on the road. The scenario in (5c) exemplifies social forces. Michelle tells Tom to say home, and, in effect, successfully opposes his tendency to go to work. Finally, the sentence in (5d) illustrates the effect of institutional forces. Here, the storm warning system for a large city brings about the closure of schools in the area.

Talmy noted that people are able to interpret interactions involving different kinds of forces. For example, consider the scenario depicted in Figure 8, in which a woman is standing in a raft and pointing in a particular direction. She indicates the direction she wants to move by pointing. In the left panel, she wants to move away from the cone while in the right panel she wants to move towards it. In both scenes, the fans turn on and push the raft to the cone. If intentions are analogous to physical forces, as assumed in the dynamics model, people should prefer to say that the fans *caused* the woman to reach the cone for the left panel while for the right panel they should prefer to say that the fans *enabled* her to reach the cone. Similarly, if the woman is pointing towards the cone but is blown away from it, people should prefer to say that she was *prevented* from reaching the cone. These predictions were tested in the next experiment.

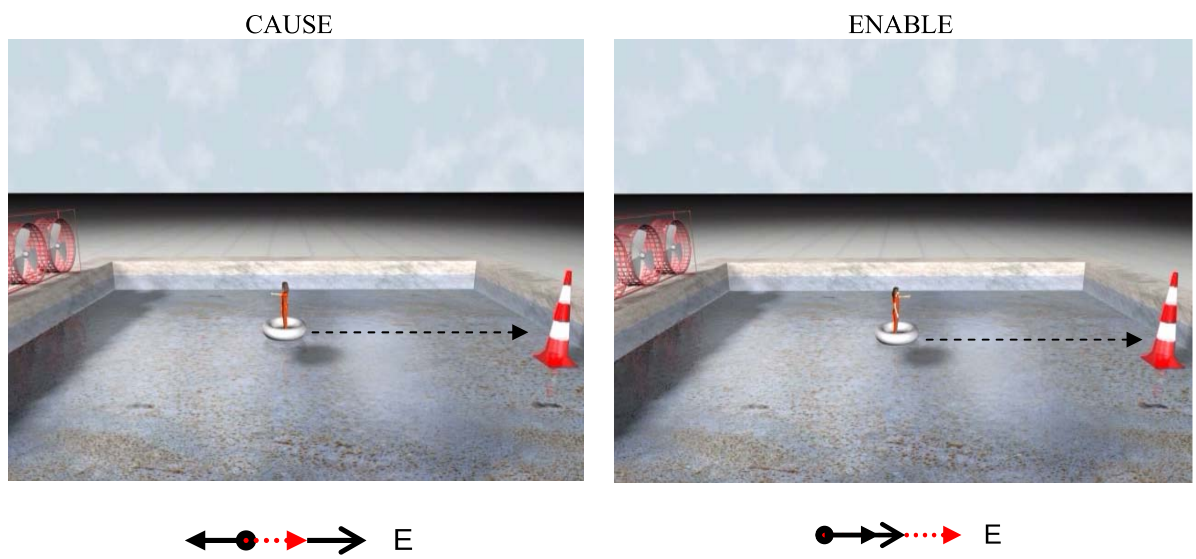


Figure 8. CAUSE (left panel) and ENABLE (right panel) animations with the patient tendency based on the woman's intention as indicated by her pointing

Experiment 5

In this experiment participants saw two types of animations. Half of the animations depicted physical causation and involved the same animations used in Experiments 1-4, namely fans acting on a boat. The remaining animations were exactly the same, except that the motorized boat was replaced with a round raft with a woman in it who pointed either towards or away from the cone (the endstate). It was predicted that people would interpret the woman's pointing as indicating patient tendency. Animations in which all of the forces were physical constituted the physical-force-only condition while animations in which the affector force was physical but the patient's force was an intention constituted the mixed-force condition.

Method

Participants. The participants were 18 University of Memphis undergraduates. All participants were native speakers of English.

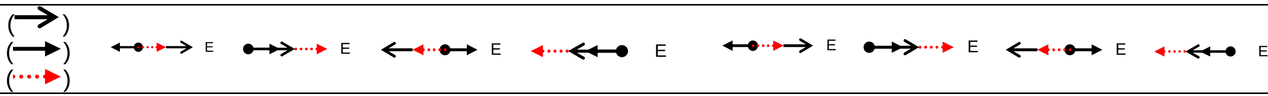
Materials. Eight 3D animations were constructed from the four, 1-dimensional base configurations depicted in Table 7. Four of the animations were the same as those used in Experiments 1-2. The remaining animations depicted a similar scene, but the boat was replaced with a woman in a raft who pointed either towards or away from the cone. Each set of animations included a CAUSE, ENABLE, PREVENT, and unspecified configuration. In the physical-force-only condition, as in previous experiments, the boat initially moved toward the center under its own power. Once it reached the center, the fans started blowing. In the mixed-force condition, there was no initial movement of the raft; instead, the tendency was indicated by the woman's pointing. In all other respects, the animations in the two conditions were the same. The animations ended when the patient hit the cone or neared the side of the pool, and the camera angle for all of the animations was the same.

The directions of the forces entered into the physics simulator are shown in Table 7. The magnitudes of these forces were as follows. For configuration 1, $||\mathbf{A_w}|| = 2N$ and $||\mathbf{P_w}|| = 1N$. For configuration 2, $||\mathbf{A_w}|| = .5N$ and $||\mathbf{P_w}|| = .5N$. For configuration 3, $||\mathbf{A_w}|| = 1.5N$ and $||\mathbf{P_w}|| = 1N$. For configuration 4, $||\mathbf{A_w}|| = .5N$ and $||\mathbf{P_w}|| = .5N$. In the remaining configurations, $\mathbf{P_w}$ was an intentional force, so the only force entered into the physics simulator was $||\mathbf{A_w}||$, which equaled 1N. The duration of the animations for configurations 1-8 were 17, 8, 6, 17, 10, 5, 10, and 5 seconds respectively.

Procedure. The procedure was similar to that in Experiments 1-3. The animations were presented in random order on Windows-based computers. After each animation, participants chose a sentence that best described the occurrence. All the sentences were structurally the same (*The fans ____ the boat to [from] hit[ting] the cone* or *The fans ____ the woman to [from] reach[ing] the cone*). The sentences differed in the main verb, which was either *causeD*, *helpED*, or *preventED*. The order of the sentence choices was presented randomly for each animation except for the last choice, which was always *none of the above*. Participants indicated their answers by clicking a radio button next to their choice.

Design. Participants saw all eight animations. There were three factors: ConfigType (CAUSE, ENABLE, PREVENT, UNSPECIFIED), ForceType (Physical, Mixed), and ResponseType (Cause, Enable, Prevent, NoVerb).

Table 7.
Experiment 5 predictions and results by configuration and response type (mean (SD))

	Physical Forces Only				Physical and Psychological Forces			
Config. #	1 N-N-Y	2 Y-Y-Y	3 Y-N-N	4 N-N-N	5 N-N-Y	6 Y-Y-Y	7 Y-N-N	8 N-N-N
Affector ( ) Patient ( ) Result. ( )	 E	E	E	E	E	E	E	E
Predicted	CAUSE	ENABLE	PREVENT	Unspecified	CAUSE	ENABLE	PREVENT	Unspecified
"Cause"	**.94** (.236)	-	-	-	**.94** (.235)	.11 (.323)	-	-
"Help"	-	**.89** (.323)	-	.06 (.236)	-	**.83** (.383)	-	.11 (.323)
"Prevent"	.06 (.236)	-	**1** (0)	.11 (.323)	-	-	**1** (0)	.11 (.323)
"No verb"	-	.11 (.323)	-	**.84** (.383)	.06 (.236)	.06 (.236)	-	**.78** (.428)

Results

The results indicated that the dynamics model can be extended to situations involving non-physical forces. The bottom of Table 7 shows the proportion of times each response type was chosen for the animations in the physical-force only and the mixed-force conditions. The results show that people treated the woman's intention as if it were a physical force.

This conclusion is supported by log-linear modeling. A log-linear model based on the factors ConfigType (4), ForceType (2), and ResponseType (4) and their two- and three-way interactions was fitted to the observed frequencies. A Pearson's chi-square indicated that such a model agreed well with the observed frequencies, and there was no evidence that the observed frequencies, $\chi^2(12, N = 144) = 2.622$, $p < .998$, differed from the predictions of the model.

Each factor and interaction was removed from this model to examine its relative contribution to the model's fit. As predicted, removing the interaction between ConfigType and ResponseType from the model resulted in a significant decrease in the fit, $\chi^2(12, N = 144) = 210.88$, $p < .0005$. This interaction indicates that responses differed across configuration types. Removal of the remaining two-way interactions, ForceType*ResponseType, $\chi^2(4, N = 144) = .34$, $p = .987$, ForceType*ConfigType, $\chi^2(3, N = 144) = .08$, $p = .994$, and the one three-way interaction, ConfigType*ResponseType*ScenarioType, $\chi^2(12, N = 144) = 2.67$, $p > .998$, did not have a significant effect on the fit of the model. Removal of the main factors of ConfigType, $\chi^2(3, N = 144) = .28$, $p = .964$, and ResponseType, $\chi^2(4, N = 144) = 2.73$, $p = .604$, did not have a significant effect on the fit of the model either. Importantly, removing the main factor of ForceType, $\chi^2(1, N = 144) = .07$, $p = .796$, had no appreciable effect on the fit of the model. The nonsignificance of this factor and its associated interactions indicates that there is no evidence that people treated the mixed-force animations differently from the physical-force-only animations. Rather, the most important factor in how people responded to the animations was how the forces were configured.

Combining the responses across ForceType, the Pearson's chi-square indicated that participants chose CAUSE descriptions over the other descriptions for the animations in which the patient did not have a tendency for the endstate, but reached it anyway (1 & 5), $\chi^2(4, N = 36) = 86.89$, $p < .0005$. Participants chose ENABLE descriptions when the patient had a tendency for the endstate and then was assisted by the fans (2 & 6, N = 36), $\chi^2(4) = 70.09$, $p < .0005$. Participants chose PREVENT descriptions when the patient had a tendency for the endstate but was blown away from it (3 & 7, N = 36), $\chi^2(4) = 91.17$, $p < .0005$. Finally, participants chose *none of the above* when the force configurations did not map onto any one of the four main kinds of configurations, $\chi^2(4, N = 36) = 60.61$, $p < .0005$.

The results for this experiment demonstrate the dynamics model's ability to explain causal judgments that require the consideration of non-physical forces. In addition, the animations depicting the CAUSE and ENABLE configurations provide further evidence against a kinematics approach to causation since the two configurations were exactly the same in terms of kinematics. The only difference was the direction in which the woman was facing.

To further test the extensional adequacy of the dynamics model, we need to know whether it can account for causal judgments in which all of the influences are non-physical. Consider, for example, the pairs of scenarios depicted in Figure 9. Each pair represents two frames[18] from an animation involving three people. The woman is the patient, the police officer is the affector, and the man on the corner is the endstate. In the first frame of the top row, the woman points away from the man to indicate that she does not wish to go toward him. In the second frame, the officer gestures that she should cross the street towards the man. She crosses the street (albeit reluctantly) toward the man. This animation involves physical motions, but what makes it a

[18] The frames shown in Figure 10 are actually cropped versions of the larger scene that is shown in Figure 11.

CAUSE scenario are the intentions and desires of the patient and affector. Specifically, it depicts causation because the woman does not want to approach the man, but the officer wants her to anyway, and the woman complies with the wishes of the officer. The dynamics model predicts that people will describe this scene with a sentence like *The officer caused the woman to walk to the man*.

The second pair of frames comes from an animation instantiating an ENABLE scenario. In this animation, the woman points towards the man, indicating that she wants to go toward him. The officer gestures for her to cross in the direction of the man, and she does so. The dynamics model predicts that people will describe this animation with an ENABLE verb, e.g., *The officer let the woman walk to the man.*

The third pair of frames comes from an animation instantiating a PREVENT configuration. In the first frame, the woman points toward the man to indicate that she wants to go to him. However, the officer gestures to the woman that she must walk in another direction, away from the man. The woman complies with the officer and so does not approach the man. According to the dynamics model, people should describe this situation with a PREVENT sentence such as *The officer prevented the woman from walking to the man*.

The fourth row shows a pair of frames from an animation instantiating a DESPITE configuration. In the first frame, the woman points toward the man to indicate that she wants to go to him. The officer gestures to the woman that she must cross the other street away from the man. However, the woman defies the officer and approaches the man. According to the dynamics model, people should describe this situation with a DESPITE sentence like *The woman crossed to the man despite the officer*.

The fifth row depicts frames from an animation instantiating an UNSPECIFIED configuration that violates the spanning restriction. In this animation, the woman points toward the man, indicating that she wants to go to him, the officer gestures for her to cross the street in that direction, but she crosses the other street. This scenario violates the spanning restriction since the observed resultant lies outside the region bounded by the intentional forces implied by the woman's and the officer's gestures. The dynamics model does not specify categories for configurations of force that violate the spanning restriction. Therefore, it predicts that people will view the scene as non-causal. These predictions were tested in the next experiment.

Experiment 6

The central question addressed in this experiment is whether the dynamics model can account for the representation of social causation. Participants were presented with situations in which all of the forces were non-physical. The affector force was indicated by the pointing gestures of a police officer and the patient force, by the gestures of a woman. In some animations, the intentions of the police officer and the woman were in conflict, while in other animations, they were in concordance. In certain animations, the woman went where she wanted to go, while in other animations she did not. According to the dynamics model, these are the basic ingredients of how people recognize and represent causation in social situations.[19]

Method

Participants. The participants were 20 Emory University undergraduates. All participants were native speakers of English.

Materials. Ten 3D animations were constructed, two each instantiating CAUSE, ENABLE, PREVENT, DESPITE, and UNSPECIFIED configurations. The five base configurations are depicted in Table 8. Two animations were made from each base configuration. In five of these animations, the endstate, the man, was located at the corner closest to the taxi, as shown in Figure 9. In the remaining animations, the endstate was located at the corner closest to the VW beetle, as shown in Figure 10.

Figure 10. Entire scene shown to participants in Experiment 6.

[19] The animations used in this experiment are taken from a separate line of research with Larry Barsalou and Aron Barbey, both of whom had a major role in their design.

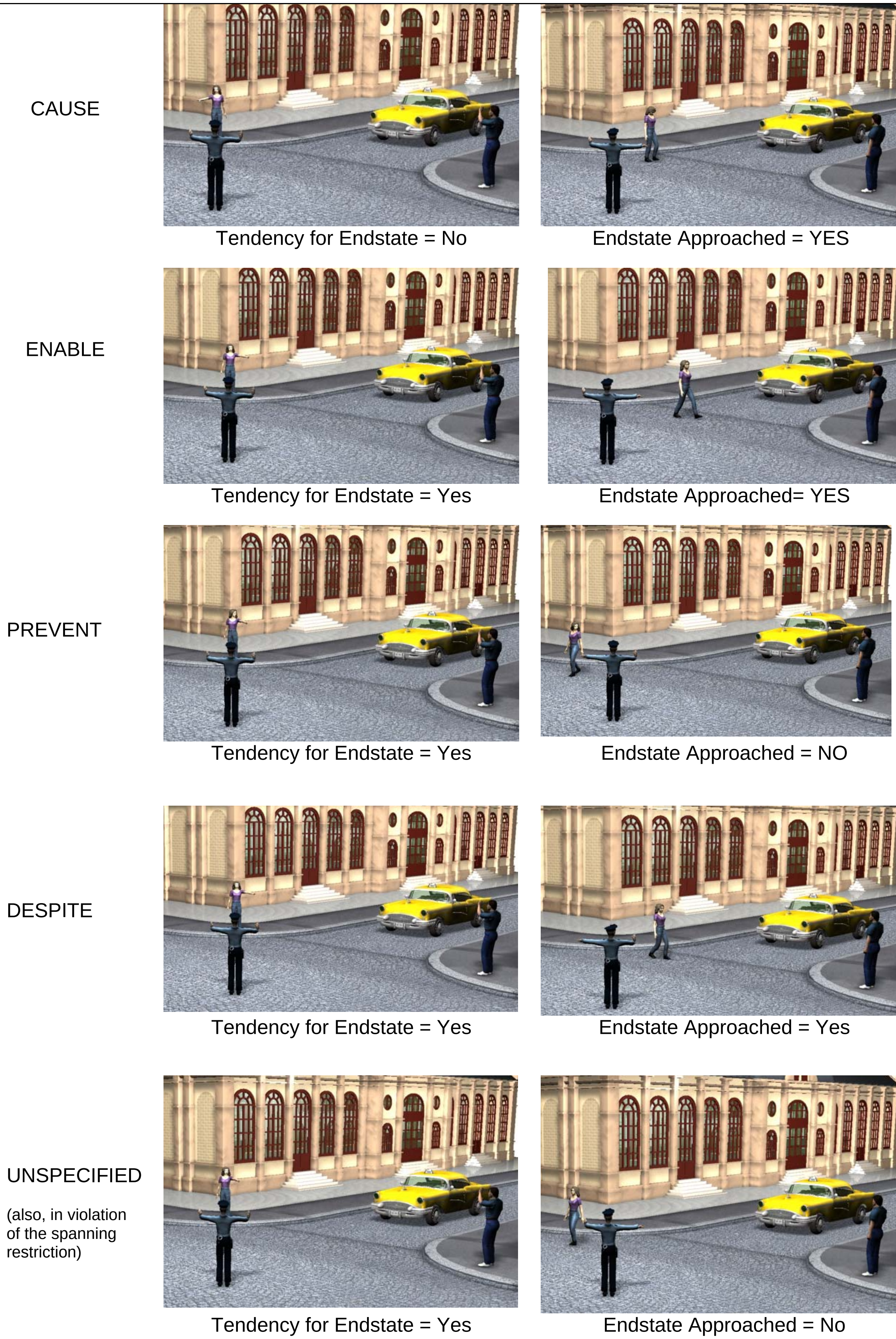


Figure 9: Sample frames from social situations instantiating CAUSE, ENABLE, PREVENT, DESPITE, and UNSPECIFIED configurations.

The animations had four main parts. In the first, the woman walked to a corner of an intersection and then stopped. As the woman walked, an officer stood in the middle of an intersection with his hand held up to imply that the cars should remain stopped. In addition, a man stood at another corner, waving at the woman to get her attention. In the second part, the woman pointed to a corner, either toward or away from the man. If the woman wanted to go to the man, she pointed in his direction and waved at him. If the woman did not want to go toward the man, she not only pointed away from him, but also avoided looking in his direction. In the third part, the officer gestured for the woman to start walking using a circular motion with one of his hands. The officer then pointed to one of the corners with the same arm he used to gesture to the woman. The officer always gestured with the arm that was closest to the corner to which he ultimately pointed. In the fourth part, the woman crossed the street. The total length of each animation was approximately 17 seconds. In each animation, the scene, camera angle, and lighting were exactly the same

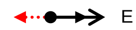
Procedure. The procedure was similar to that in Experiments 1, 2, 3, and 5. The animations were presented in random order on Windows-based computers. After each animation, participants chose a sentence that best described the occurrence. Three of the choices were based on the exact same sentence (*The officer ____ the woman to[from] walk[ing] up to the man*), except for the main verb which was either *caused*, *enabled*, or *prevented*. The DESPITE option was the sentence *The woman walked to the man despite the officer*. The last option was *none of the above*. The order of the sentence choices was changed randomly for each animation except for the last choice, which was always the option *none of the above*. Participants indicated their answers by clicking a radio button next to their choice.

Design. Participants saw all ten animations. There were two factors: ConfigType (CAUSE, ENABLE, PREVENT, DESPITE, UNSPECIFIED) and ResponseType (Cause, Enable, Prevent, Despite, NoVerb).

Results

The results indicate that the dynamics model can be extended to explain people's judgments about social causation. The bottom of Table 8 shows the proportion of times people chose each of the five possible response types (CAUSE, ENABLE, PREVENT, DESPITE, NOVERB) to describe each of the configuration types. The results provide further evidence that the dynamics model is able to distinguish CAUSE from other causal concepts, and that causal relations can be

Table 8. Experiment 6 predictions and results by configuration and response type (mean (SD))

Configuration #	1	2	3	4	5
Type	CAUSE N-N-Y	ENABLE Y-Y-Y	PREVENT Y-N-N	DESPITE Y-N-Y	UNSPECIFIED Y-Y-N
Affector ( ) Patient ( ) Result. ( )					
"Cause"	**.84** (.239)	-	-	-	-
"Enable"	.13 (.226)	**.94** (.158)	-	.08 (.187)	.16 (.336)
"Prevent"	.03 (.115)	-	**.90** (.315)	-	.08 (.187)
"Despite"	-	.03 (.115)	-	**.92** (.187)	-
"No verb"	-	.03 (.115)	.10 (.315)	-	**.76** (.348)

identified from a single occurrence. In addition, the results support the claim that the concept of CAUSE is based on the relationship between several forces. If, for example, people considered only the affector force without factoring in that of the patient (i.e., the tendency of the woman), there would be no way of distinguishing between the concepts of CAUSE and ENABLE.

The above conclusions are supported by log-linear modeling. A log-linear model based on the factors ConfigType (5), and ResponseType (5) and a single two-way interaction was fitted to the observed frequencies. A Pearson's chi-square indicated that such a model agreed well with the observed frequencies as there was no evidence for a difference between the predicted frequencies and the actual frequencies, $\chi^2(4, N = 200) = 1.66$, $p < .798$.

Each factor and interaction was removed from this model to examine its relative contribution to the model's fit. As predicted, removing the interaction between ConfigType and ResponseType from the model resulted in a significant decrease in the fit, $\chi^2(16, N = 200) = 367.01$, $p < .0005$. This interaction indicates that responses differed across configuration types. Removal of the main factors of ConfigType, $\chi^2(4, N = 200) = 1.59$, $p = .812$, and ResponseType, $\chi^2(4, N = 200) = 3.33$, $p = .505$, did not have a significant effect on the fit of the model.

Focusing on particular response types, Pearson's chi-square indicated that participants chose *CAUSE* descriptions more often than the other types of descriptions for the animations in which the woman did not have a tendency for the endstate, but ended up walking toward the man (1), $\chi^2(4, N = 40) = 98.05$, $p < .0005$. Participants chose *ENABLE* descriptions when the woman had a tendency for the endstate and was directed toward the endstate by the officer (2), $\chi^2(4, N = 40) = 126.44$, $p < .0005$. Participants chose *PREVENT* descriptions when the woman had a tendency for the endstate but was directed away by the officer (3), $\chi^2(4, N = 40) = 79.73$, $p < .0005$.

Participants chose *DESPITE* descriptions when the woman had a tendency for the endstate and walked toward it in opposition to the direction of the officer (4), $\chi^2(4, N = 40) = 79.62$, $p < .0005$. Finally, participants chose "none of the above" when the force configurations did not map onto any one of the four main kinds of configurations (5), $\chi^2(4, N = 40) = 106$, $p < .0005$.

Participants were quite willing to use the verb *enable* to describe the ENABLE configurations in the present experiment. As discussed earlier, the various ENABLE verbs differ in what they imply about what might occur in the absence of the affector. The verb *help* (and sometimes *enable*) leaves open the possibility that the result could occur in the absence of the affector. In contrast, the verbs *allow, let*, *permit* (and sometimes *enable*) imply that the result could not occur without the force of the affector. Participants' willingness to use the verb *enable* indicates that the ENABLE configuration of forces is not restricted to the verb *help*. However, what may have made the ENABLE animations in this experiment particularly conducive to the verb *enable* (or *allow*) is that the police officer, in effect, removed an institutional force acting against the woman's crossing the street. The officer's gestures in the ENABLE animation indicated that he was removing the prohibition against crossing a street in busy traffic, hence allowing the woman's tendency to be realized. As discussed in Appendix A, it may be possible to represent the removal of a force by interpreting a configuration in terms of the inverse of the affector vector (~**A** instead of **A**) and/or in terms of preventing a prevention. Finally, it is interesting to note that participants were somewhat more willing to use the verb *enable* to describe the CAUSE animation (13%) than in previous experiments. This might reflect a greater willingness to use the verb *enable* to describe social causation than physical causation.

The current experiment demonstrated that the spanning restriction is enforced even in situations involving non-physical forces. In the case of the UNSPECIFIED configuration in the current experiment, the woman had a tendency for the endstate, the officer directed her to the endstate, but she walked away. This set of factors does not add up to a causal situation, as reflected in participants' preference for the choice *none of the above* to describe this situation. We may encounter such situations in real life and engage in a similar kind of reasoning. For example, suppose that Jane wanted to go to the movies (but didn't have means to get there) and that her friends invited her to go with them, but then she didn't go. A person observing this sequence of events might be puzzled by Jane's behavior or come up with possible explanations, e.g., Jane forgot about the appointment, or she changed her mind. However, regardless of the explanation, the interaction between Jane and her friends cannot be said to instantiate a causal or preventative relation. The dynamics model explains why.

Finally, the findings provide further evidence against dependency models, which are unable to predict that causation across different domains should share anything more than statistical or counterfactual dependencies. It is striking, then, that in these experiments, the distinctions used in judgments about physical causation were the same as those used for social causation. Across domains, causal judgments reflected attention to the dimensions of tendency, concordance, and result. Dependency models—unlike the dynamics approach—cannot motivate why these similarities should exist.

General Discussion

In order to understand how people learn and reason about causal relationships, we need to understand how causal relations are represented in the mind. A theory of causal representation should be able to pick out the range of situations that people consider to be causal while excluding situations that people do not consider to be causal. Dependency models—both probabilistic and counterfactual—fail in this respect because the range of situations they classify as causal is too broad. In addition, a theory of causal representation should explain people's ability to determine causal relationships on the basis of a single observation. Here, again, dependency models have problems because they hold that people require multiple observations to establish causation. Some have suggested that the identification of causal relationships on the basis of a single observation might be accomplished by the application of causal categories formed on the basis of covariational information. However, causal categories based on dependency information would still be too inclusive. Moreover, if causal categories depended on pre-stored causal categories, people would only be able to identify causal relations that conformed to their prior experience, which raises the developmental question of how such an account would ever get off the ground. Physicalist models offer an account of how people might identify causal relations on the basis of a single example, but in the past, they have not been able to distinguish causation from other cause-related concepts such as enablement and prevention. The dynamics model, introduced in this paper, is able to both differentiate causation from non-causation and explain how people identify causal relations from a single observation. According to the dynamics model, causal concepts are represented in terms of relationships among various forces and a position vector.

The model was supported by a series of experiments in which participants watched 3D animations of complex causal events. The animations were the same in every respect except for the underlying configurations of forces, which produced different patterns of motion, and so instantiated different causal relationships. In Experiment 1, people distinguished different causal interactions occurring within a single dimension. In Experiment 2, people distinguished causal interactions occurring over more than one dimension. Experiment 3 showed that the results from Experiments 1 and 2 were generalizable to other verbs and scenarios. It also provided support for the proposed representation of DESPITE, another causal concept predicted by the model. Experiment 4 provided evidence against the possibility that people's causal judgments in Experiments 1-3 were based only on their kinematics by showing that people are sensitive to violations of the underlying dynamic properties of an event. Experiment 5 demonstrated that the dynamics model could be applied to situations involving both physical and intentional forces. Finally, Experiment 6 showed that the dynamics model makes accurate predictions regarding people's causal judgments of purely social interactions. The results from this last experiment show that the dynamics model extends beyond physical causation.

Patterns of force underlying counterfactual thinking and probabilistic causes

The dynamics model is intended as an account of the core concept of causation. While it has certain advantages over dependency models, it is not necessarily incompatible with them to the extent that they can be viewed as tests of causation. Interestingly, the dynamics model can be used to model at least certain kinds of counterfactual judgments and probability distributions.

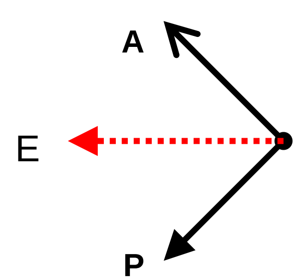


Figure 11: A configuration of forces associated with CAUSE.

Consider, for example, the configuration of forces in Figure 11, which was used in Experiment 2. People's overwhelming interpretation of this configuration was that the affector

(**A**) caused the patient to reach the endstate (E). This configuration can be used as the basis of a counterfactual. As discussed earlier, a counterfactual criterion of causation holds that an event *c* is a cause of an event *e* if and only if it is the case that if *c* had not occurred, *e* would not have occurred. In Figure 11, if the affector force were removed, the resultant would not point towards the endstate, supporting the conclusion that the affector is a cause because it is necessary for the result. In this way, the dynamics model specifies the knowledge needed to conduct counterfactual reasoning about causation.

The model can also motivate why causation is associated with a positive statistical dependency. Imagine a group of configurations instead of just one configuration. The proportion—hence probability— of configurations in which the resultant points towards the endstate to produce an effect could be determined with respect to the presence of the affector, P(E|A), and its absence, P(E|~A). If most of the configurations were CAUSE configurations, P(E|A) would be greater than P(E|~A), implying a positive correlation. If most of the configurations were PREVENT configurations, P(E|A) would be less than P(E|~A), implying a negative correlation. This link between configurations and correlations explains why statistical dependency is often a valuable cue to causation.

In the case of an individual occurrence, causal relationships are deterministic. However, in the case of multiple instances, causal relationships can be probabilistic, as when we say *Heavy rains cause flooding*. However, this does not require that the underlying representation change from vectors to probabilities. In the dynamics model, uncertainty is built into the representation of causation because people do not know the exact magnitude of the vectors. Over chains of configurations, in which the resultants of one configuration are used as the affectors in the next, these variations in the magnitude of the vectors can lead to different conclusions even though the component configuration are the same. This means that if we observed multiple chains, the overall conclusion would be probabilistic. The dynamics model is compatible, then, with both deterministic and probabilistic causation (see Barbey & Wolff, 2006).

Other dependency models

Besides the dependency models discussed in the introduction, two other dependency models have had an important impact on research on the representation of causation. According to associative learning models, causal relations are defined in terms of association weights (Baker, Murphy, & Vallée-Tourangeau, 1996; Shanks & Dickinson, 1987). One strength of these models is that they provide an account of how causal relations might be determined through an associative learning process as defined by, for example, the *Rescorla-Wagner* learning rule. However, in associative learning models, causal categories are typically represented in terms of a value on a single output that can be either positive, negative, or zero. Consequently, these accounts are basically limited to the distinction between generative and preventative causation (Cheng & Holyoak, 1995; Cheng, Park, Yarlas, & Holyoak, 1996; Cheng, 1997). These models also assume learning procedures that require multiple observations in order for causal relations to be established, leaving unexplained people's ability to discern a causal relation from a single observation except when such an observation matches an already established causal relationship.

Another influential dependency theory of causation is Goldvarg and Johnson-Laird's (2001) model theory. According to the model theory, causal relations are intrinsically modal, that is "[t]hey are not merely about what occurred but also about what might have occurred" (p. 576). In this theory, various causal-related concepts, including CAUSE and ALLOW,[20] are differentiated from each other in terms of possible co-occurrences of the cause and the effect. At its core, the model theory defines causal concepts in terms of necessity and sufficiency. The claim *A caused B* is false if A can occur without the occurrence of B, that is, if A is not a *sufficient* condition for the occurrence of B. The claim *A allowed B* is false if B can occur without A, that is, the claim is false if A is not a *necessary* condition for the occurrence of B. In sum, causes are factors that are sufficient, and maybe also necessary, for their effects while allowers are factors that are necessary, but not sufficient, for their effects (see Mandel, 2003).

The model theory has several strengths. First, it proposes how causal relations might be determined on the basis of a single example (but see Wolff & Song, 2003). Second, it is supported by Mandel and Lehman's (1998) finding that people tend to define causation (and prevention) in terms of tests of sufficiency more than in terms of tests of necessity. Another important contribution is that it addresses the problem of how to distinguish the concept of CAUSE from ALLOW (or ENABLE).

However, certain predictions of the model theory are not supported by the results in Experiments 1-6. In Experiment 2, the affector force in the CAUSE configuration was an

---

[20] Goldvarg and Johnson-Laird prefer the mnemonic *ALLOW* to the mnemonic *ENABLE* because they suggest that the verb *allow* is more neutral with respect to intentionality than is the verb *enable*.

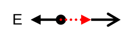
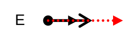
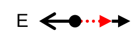
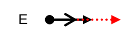
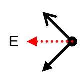
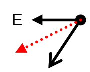
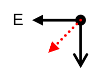
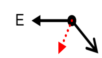
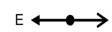
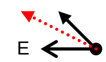
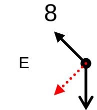
necessary but not sufficient condition for the result to occur. The model theory predicts people should have described these configurations as instances of ALLOW when, in fact, they were described as instances of CAUSE. In Experiment 5, the affector in the ENABLE animation involving the woman in the raft was a necessary and sufficient condition for the result. The model theory predicts people should have viewed this scene as causal when, in fact, they viewed it as enabling. The model theory addresses many of the shortcomings of other dependency models, but its semantics is based on necessary and sufficient conditions, which often lead it to make predictions that do not match people's judgments of causation.

Implications for physicalist models of causation

The results from Experiments 1-6 not only highlight problems for dependency models, they also reveal some of the limitations of prior physicalist models. In the psychological literature, causation has been characterized as a transmission of motion (e.g., Michotte, 1946/1963; Kruschke & Fragassi, 1996) or of causal impetus (Bullock, Gelman & Baillargeon, 1982; Hubbard & Ruppel, 2002; Shultz, 1982). Similarly, in the philosophy literature, causation has been reduced to a transfer of a conserved quantity, such as momentum or energy (Aronson, 1971; Fair, 1979; Salmon, 1994, 1998). The experiments in this paper allow for a fairly direct test of these proposals. Because the materials were generated from a physics simulator, the amount of energy and momentum transferred from the cause to the effect can be calculated. If causation is reducible to the transmission of energy or momentum, then it should be possible to identify causal relationships on the basis of these quantities.

Table 9 shows the amounts of momentum and energy involved in the boat scenarios from Experiments 1 and 2. I explain how these quantities were calculated in Appendix C. The results from this analysis indicate no relationship between the amounts of momentum and energy transferred and various causal relationships. For example, in Experiment 2 (configuration #1), the CAUSE animation involved positive energy transfer, but in Experiment 1 (configuration #1), the CAUSE animation involved negative energy transfer. The transfer of energy was negative because the fans decreased the velocity of the boat and, hence, the boat's kinetic energy. Similarly, PREVENT animations were sometimes associated with negative energy transfer (Exp. 1, config 4; Exp. 2, config. 5 and 6) and other times with positive energy transfer (Exp. 2, config. 3 and 4). PREVENT involved positive energy transfer when, for example, the fans sped the boat past the endstate (Exp. 2, config. 3 and 4). ENABLE animations were consistently associated

Table 9. Amount of energy and momentum transferred from the fans to the boat in the animations used in Experiments 1 and 2[21]

				Experiment 1				
Config. #	1	2	3	4	5	6	7	8
Affector (→) Patient (→) Result. (→)								
Type	CAUSE	ENABLE	ENABLE	PREVENT	DESPITE	Unspecified	Unspecified	Unspecified
Energy transferred (joules)	-.060	.727	.509	-.060	-.277	.727	-.277	.509
Momentum transferred (kg*m/s)	.817 (at 180°)	.817 (at 0°)	.481 (at 0°)	.817 (at 180°)	.481 (at 180°)	.817 (at 0°)	.481 (at 180°)	.481 (at 0°)
				Experiment 2				
Config. #	1	2	3	4	5	6	7	8
								
Type	CAUSE	ENABLE	PREVENT	PREVENT	PREVENT	PREVENT	Unspecified	Unspecified
Energy transferred (joules)	.118	.348	.269	.118	-.005	-.040	.269	-.005
Momentum transferred (kg*m/s)	.451 (at 39.5°)	.521 (at 0°)	.488 (at 18.5°)	.451 (at 39.5°)	.424 (at 59.5°)	.115 (at 180°)	.488 (at 18.5°)	.424 (at 59.5°)

with positive energy transfer, and the DESPITE animation was associated with negative energy transfer. However, since CAUSE and PREVENT animations were associated with both negative and positive energy transfer, energy transfer cannot be used to identify these relations. Transfer of momentum was no better a diagnostic. In Experiments 1 and 2, for example, CAUSE and PREVENT were exactly the same in terms of transfer of momentum, and in Experiment 2, transfer of momentum in one of the ENABLE animations matched the transfer of momentum in one of the unspecified configurations.

Dowe's Conserved Quantity Theory (2000) proposes that causation does not depend on transfer or transmission, but rather the exchange of conserved quantities, such as energy, momentum, or charge. As Dowe argues, the notion of exchange is weaker than that of transfer. In particular, CQ theory does not require that the energy or momentum move from the cause to the effect. The results in Table 9 support Dowe's CQ theory, but also reveal its limitations. Every transfer of a conserved quantity involves an exchange of conserved quantities, and vice versa.

[21] In the interest of space, configurations 9 and 10 from Experiment 2 are not included in Table 9.

For example, when the fans (i.e., the air molecules) transferred energy to the boat, the boat extracted energy from the air molecules. Thus, all of the interactions in Experiments 1 and 2 involved an exchange of conserved quantities. Because all of the interactions involved exchanges of energy, CQ theory cannot tell us which of the interactions were causal, as opposed to enabling, preventative, or unclassifiable.

Causation and time

In most theories of causation, causation supervenes on time. For example, in probability distribution models, causation is reduced to probabilities of events, i.e., segments of time. Time is also crucial in the way causation is characterized in the linguistics literature, where causal relations are viewed as necessarily composed of two events—a causing subevent and a resulting subevent—that occur in sequence (Dowty, 1979; Croft, 1991; Jackendoff, 1990; Levin & Rapoport, 1988; Levin & Rappaport Hovav, 1995; Pustejovsky, 1991; Van Valin, 1990, among others). However, some have suggested that the relationship between causation and time might be the other way around, specifically, that people might individuate events in terms of causation (Davidson, 1969/1980; Bullock et al., 1982).

The proposal that causation individuates events is circular if causation is itself composed of events (Avrahami & Kareev, 1994). However, in force dynamics, causation does not depend on events; rather, it depends on space. From this perspective, it may be possible to re-evaluate the proposal that events might unitized in terms of causation.

From a force dynamic perspective, causation is not tied to a sequence of two events, though it can certainly be instantiated by such sequences. Indeed, the dynamics model provides a new explanation for why billiard-ball events, like the ones studied by Michotte (1946/1963), are construed as causal (see also Leslie, 1994). Specifically, when object A hits object B, it exerts a force on object B that opposes B's tendency to remain at rest due to friction. When the forces acting on object B are added together, they sum to a resultant that accelerates object B. Many of Michotte's findings can be motivated by the dynamics model. For example, spatial contiguity is important since a configuration of contact forces requires physical contact. Temporal contiguity is also important because the configuration of forces that results from this contact lasts for only a moment. In effect, the dynamics model shows how launch events can be linked to other kinds of causation (e.g., the causal situations examined in this paper) and need not be viewed as the product of an innate perceptual mechanism (for a review, see Scholl & Tremoulet, 2000).

One of the most interesting consequences of force dynamics is that it can be applied to statics, that is, to situations in which nothing happens, and yet there is a continuing state of causation, as in *Dirt caused the valve to stay open* or *Tiny barbs on the stinger cause it to remain in the wound*. Force dynamics can also be applied to situations where there is continuous change without a clear temporal separation of cause and effect events, as in *Gravity causes the earth to orbit the sun* or *Greenhouse gases are causing temperatures to rise.* The dynamics approach accounts naturally for these types of causal relationships because it does not require that cause and effect events occur in succession. Rather, what is required is simultaneity. Causal relations can hold for a single moment or an indefinite period of time. Even in the case of collision events, in which the actions of the causer clearly precede changes in the patient, there is a single moment—the moment of contact—in which there is a convergence of forces. According to the dynamics model, it is this moment in time that is critical to defining such interactions as causal. From a force dynamic perspective, temporal priority is not a requirement of causation; rather, it is an artifact of the way forces often converge.

Conclusions

Causation is an atemporal spatial arrangement of forces. The implications of this conclusion are in sharp contrast to those assumed in dependency models. In probability distribution models, the properties of a causal event matter very little to the way causal events are represented. All that is required is that such events be countable. Thus, in these theories, causation is largely a product of the mind, namely, its ability to keep track of event frequencies and correlations. According to the dynamics model, in contrast, causation is mostly a product of the world, and representing causation involves representing the physical quantities that bring these causal events about. Not only can causes be counted, they can also be sensed and perceived, and what we feel can factor directly into how causes are represented, namely, as patterns of forces.

References

Ahn, W., & Bailenson, J. (1996). Mechanism-based explanations of causal attribution: An explanation of conjunction and discounting effect. *Cognitive Psychology, 31*, 82-123.

Ahn, W., & Kalish, C. W. (2000). The role of mechanism beliefs in causal reasoning. In F. C. Keil & R. A. Wilson (Eds.), *Explanation and cognition* (pp. 199-225). Cambridge, MA: The MIT Press.

Ahn, W., Kalish, C. W., Medin, D. L., & Gelman, S. A. (1995). The role of covariation versus mechanism information in causal attribution. *Cognition, 54*, 299-352.

Anderson, J. M. (1971). *The grammar of case: Towards a localistic theory*. Cambridge: Cambridge University Press.

Anscombe, G. E. M. (1971). *Causality and Determination*. Cambridge: Cambridge University Press.

Aronson, J. L. (1971). On the grammar of 'CAUSE'. *Synthese, 22*, 414-430.

Avrahami, J., & Kareev, Y. (1994). The emergence of events. *Cognition, 53*, 239-261.

Baker, A. G., Murphy, R. A., & Vallée-Tourangeau, F. (1996). Associative and normative accounts of causal induction: Reacting to versus understanding a cause. In D. R. Shanks, K. J. Holyoak & D. L. Medin (Eds.), *The psychology of learning and motivation: Vol. 34. Causal learning* (pp. 1-46). San Diego, CA: Academic Press.

Barbey, A., & Wolff, P. (2006). Causal reasoning from forces. *Proceedings of the 28th Annual Conference of the Cognitive Science Society*, Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.

Barsalou, L.W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22,* 577 660.

Bigelow, J., & Pargetter, R. (1990). Metaphysics of causation. *Erkenntnis, 33*, 89-119.

Bigelow, J., Ellis, B., & Pargetter, R. (1988). Forces. *Philosophy of Science, 55*, 614-630.

Brown, D., & Clement, J. (1989). Overcoming misconceptions via analogical reasoning. *Instructional Science, 18*, 237–261.

Buehner, M. J., & Cheng, P. W. (1997). Causal induction: The Power PC Theory versus the Rescorla-Wagner Model. In M. G. Shafto & P. Langley (Eds.), *Proceedings of the nineteenth annual conference of the Cognitive Science Society* (pp. 55-60). Hillsdale, NJ: Lawrence Erlbaum Associates.

Bullock, M., Gelman, R., & Baillargeon, R. (1982). The development of causal reasoning. In W. Friedman (Ed.), *The developmental psychology of time* (pp. 209-255). London: Academic Press.

Bunge, M. (1959). *Causality: the place of the causal principle in modern science*. Cambridge, MA, Harvard University Press.

Carter, R. J. (1976). Some constraints on possible words. *Semantikos, 1*, 27-66.

Cheng, P. W. (1993). Separating causal laws from casual facts: Pressing the limits of statistical relevance. In D. L. Medin (Ed.), *The psychology of learning and motivation* (Vol. 30, pp. 215-264). New York: Academic Press.

Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review, 104*, 367-405.

Cheng, P. W. (2000). Causality in the mind: Estimating contextual and conjunctive causal power. In F. C. Keil & A. W. Robert (Eds.), *Explanation and cognition* (pp. 227-253). Cambridge, MA: The MIT Press.

Cheng, P. W., & Novick, L. R. (1991). Causes versus enabling conditions. *Cognition, 40*, 83-120.

Cheng, P. W., & Novick, L. R. (1992). Covariation in natural causal induction. *Psychological Review, 99*, 365-382.

Cheng, P. W. & Holyoak, K. J. (1995). Complex adaptive systems as intuitive statisticians: Causality, contingency, and prediction. In H. L. Roitblat & J. –A. Meyer (Eds.), Comparative approaches to cognitive science (pp. 271-302). Cambridge, MA: MIT Press.

Cheng, P. W., Park, J., Yarlas, A. S., & Holyoak, K. J. (1996). A causal-power theory of focal sets. In D. R. Shanks, K. J. Holyoak, & D. L. Medin (Eds.), *The psychology of learning and motivation: Vol. 34. Causal learning* (pp. 313-357). San Diego, CA: Academic Press.

Clement, J. (1983). A conceptual model discussed by Galileo and used intuitively by physics students. In D. Gentner & A. L. Stevens (Eds.), *Mental models* (pp. 299-324). Hillsdale, NJ: Lawrence Erlbaum Assoc.

Cohen, L. B., Amsel, G., Redford, M. A., & Casasola, M. (1998). The development of infant causal perception. In A. Slater (ed.), *Perceptual development: Visual, auditory, and speech perception in infancy* (pp. 167-209). East Sussex, UK: Psychology Press Ltd.

Copley, B. (in press). Ordering and reasoning. *The MIT Working Papers in Linguistics*.

Croft, W. A. (1991). *Syntactic categories and grammatical relations*. Chicago: University of Chicago Press.

Davidson, D. (1969/1980). “The individuation of events.” In *Essays on actions and events* (pp. 163-80). Oxford: Clarendon Press.

diSessa, A. (1993). Towards an epistemology of physics. *Cognition and Instruction, 10*, 105-225.

diSessa, A., Gillespie, N. M., & Esterly, J. B. (2004). Coherence versus fragmentation in the development of the concept of force. *Cognitive Science, 28*, 843-900.

Dowty, D. R. (1979). *Word meaning and Montague grammar*. Dordrecht: Reidel.

Dowe, P. (2000). *Physical causation*. Cambridge University Press: Cambridge, UK.

Einhorn, H. J., & Hogarth, R. M. (1986). Judging probable cause. *Psychological Bulletin, 99*, 3-19.

Fair, D. (1979). “Causation and the flow of energy.” *Erkenntnis 14*, 219-250.

Gentner, D., & Gentner, D. R. (1983). Flowing waters or teeming crowds: Mental models of electricity. In D. Gentner & A. L. Stevens (Eds.), *Mental models* (pp. 99-129). Hillsdale, NJ: Lawrence Erlbaum Associates.

Gentner, D., Holyoak, K.J., Kokinov, B. (Eds.) (2001). *The analogical mind: Perspectives from cognitive science*. Cambridge, MA, MIT Press.

Gilden, D. L. (1991). On the origins of dynamical awareness. *Psychological Review, 98*, 554-568.

Glymour, C. (2001). *The mind's arrows*. Cambridge, MA: The MIT Press.

Goldstone, R., & Barsalou, L. W. (1998). Reuniting cognition and perception: The perceptual bases of rules *and* similarity. *Cognition, 65*, 231-62.

Goldvarg, E., & Johnson-Laird, P. (2001). Naive causality: A mental model theory of causal meaning and reasoning. *Cognitive Science, 25*, 565-610.

Gopnik, A., Glymour, C., Sobel, D., Shulz, L., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: Causal maps and Bayes nets. *Psychological Review, 111*, 1-31.

Griffiths, T. L., & Tenenbaum, J. B. (in press). Elemental causal induction. *Cognitive Psychology*.

Fodor, J. (1970). Three reasons for not deriving "kill" from "cause to die." *Linguistic Inquiry 1*, 429-38.

Hagmayer, Y., & Waldmann, M. R. (2000). Simulating causal models: The way to structural sensitivity. In L. R. Gleitman & A. K. Joshi (Eds.), *Proceedings of the Twenty-Second Annual Conference of the Cognitive Science Society* (pp. 214-219). Mahwah, NJ: Erlbaum.

Hagmayer, Y., Sloman, S. A., Lagnado, D. A., & Waldmann, M. R. (in press). Causal reasoning through intervention. In A. Gopnik & L. Schulz (Eds.), *Causal learning: Psychology, philosophy, and computation.* Oxford: Oxford University Press.

Hart, H. L. & Honoré, A. (1985). *Causation in the law*. Oxford: Clarendon Press.

Hecht, H. (1996). Heuristics and invariants in dynamic event perception: Immunized concepts or nonstatements? *Psychonomic Bulletin & Review, 3*, 61-70.

Hubbard, T. L., & Ruppel, S. E. (2002). A possible role naïve impetus in Michotte's "launching effect": Evidence from representational momentum, *Visual Cognition, 9*, 153-176.

Jackendoff, R. (1983). Semantics and cognition. Cambridge, MA: The MIT Press.

Jackendoff, R. (1990). *Semantic structures*. Cambridge, MA: The MIT Press.

Joskowicz, L., & Sacks, E. (1991). Computational kinematics. *Artificial Intelligence, 51*, 381-416.

Kahneman, D., & Tversky, A. (1982). The simulation heuristic. In D. Kahneman, P. Slovic & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.

Kaiser, M. K., Proffitt, D. R., Whelan, S. M., & Hecht, H. (1992). The influence of animation on dynamical judgments. *Journal of Experimental Psychology: Human Perception and Performance, 18*, 669-690.

Kruschke, J. K., & Fragassi, M. M. (1996). The perception of causality: Feature binding in interacting objects. In *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society* (pp. 441-446). Hillsdale, NJ: Erlbaum.

Lagnado, D. A., Waldmann, M. R., Hagmayer, Y., & Sloman, S. A. (in press). Beyond covariation: Cues to causal structure. In A. Gopnik & L. Schulz (Eds.), *Causal learning: Psychology, philosophy, and computation.* Oxford: Oxford University Press.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.

Langacker, R. W. (1986). An introduction to cognitive grammar. *Cognitive Science, 10*, 1-40.

Leslie, A.M. (1994). ToMM, ToBy, and agency: core architecture and domain specificity. In L. Hirschfield, & S. Gelman (eds), *Mapping the Mind: Domain Specificity in Cognition and Culture* (119–148). Cambridge University Press.

Levin, B., & Rapoport, T. R. (1988). Lexical subordination. In *Proceedings of the 24th annual meeting of the Chicago Linguistic Society* (pp. 275-289).

Levin, B., & Rappaport Hovav, M. (1994). A preliminary analysis of causative verbs in English. *Lingua: International Review of General Linguistics 92*, 35-77.

Levin, B., & Rappaport Hovav, M. (1995). *Unaccusativity: At the syntax-lexical semantics interface*. Cambridge, MA: The MIT Press.

Levin, B., & Rappaport Hovav, M. (2005). *Argument realization*: *Research surveys in linguistics series*. Cambridge, UK: Cambridge University Press.

Lewis, D. (1973). Causation. *Journal of Philosophy, 70*, 556–567.

Lien, Y., & Cheng, P. (2000). Distinguishing genuine from spurious causes: A coherence hypothesis. *Cognitive Psychology 40*, 87-137.

Lober, K., & Shanks, D. R. (2000). Is causal induction based on causal power? Critique of Cheng (1997). *Psychological Review, 107*, 195-212.

Lombard, L. M. (1990). Causes, enablers, and the counterfactual analysis. *Philosophical Studies, 59*, 195-211.

Luhmann, C. C., & Ahn, W. (2005). The meaning and computation of causal power: Comment on Cheng (1997) and Novick and Cheng (2004). *Psychological Review, 112*, 685-692.

Mackie, J. L. (1974). *The cement of the universe*. Oxford: Oxford University Press.

Mandel, D. R. (2003). Judgment dissociation theory: An analysis of differences in causal, counterfactual, and covariational reasoning. *Journal of Experimental Psychology: General, 132*, 419-434.

Mandel, D. R., & Lehman, D. R. (1996). Counterfactual thinking and ascriptions of cause and preventability. *Journal of Personality and Social Psychology, 71*, 450–463.

Mandel, D. R., & Lehman, D. R. (1998). Integration of contingency information in judgments of cause, covariation, and probability. *Journal of Experimental Psychology: General, 127*, 269-285.

McCloskey, M. (1983). Naïve theories of motion. In D. Gentner & A. L. Stevens (Eds.), *Mental models* (pp. 299-324). Hillsdale, NJ: Lawrence Erlbaum Associates.

McCloskey, M., & Kohl, D. (1983). Naïve physics: The curvilinear impetus principle and its role in interactions with moving objects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9*, 146-156.

McCloskey, M., Washburn, A., & Felch, L. (1983). Intuitive physics: The straight-down belief and its origin. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9*, 636-649.

Michotte, A. E. (1946/1963). *The perception of causality*. New York: Basic Books.

Michotte A. E., & Thinés, G. (1963/1991). La causalité perceptive. J. Psych. Norm. Path., 60, 9 – 36, 1963. Translated and reprinted in Thinés, G., Costal, A., & Butterworth, G. (eds.), *Michotte's experimental phenomenology of perception* (pp. 66 – 87). Hillsdale, NJ: Erlbaum, 1991.

Miller, G. A., & Johnson-Laird, P. N. (1976). *Language and perception.* Cambridge, MA: Harvard University Press.

Norman, D. A., Rumelhart, D. E., & the LNR Research Group. (1975). *Explorations in cognition*. San Francisco: W. H. Freeman.

Oakes, L. M. (1994). The development of infants' use of continuity cues in their perception of causality. *Developmental Psychology 30*, 869-79.

Pearl, J. (2000). *Causality: Models, reasoning, and inference*. Cambridge, UK: Cambridge University Press.

Pinker, S. (1989). *Learnability and cognition: The acquisition of argument structure*. Cambridge, MA: The MIT Press.

Proffitt, D. R., & Gilden, D. L. (1989). Understanding natural dynamics. *Journal of Experimental Psychology: Human Perception and Performance, 15*, 384-393.

Pustejovsky, J. (1991). The syntax of event structure. *Cognition, 41*, 47-81.

Robertson, D. A., & Glenberg, A. M. (1998). Force dynamics in language and cognition. In *Proceedings of the 20th annual meeting of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum Associates.

Runeson, S., & Frykholm, G. (1983). Kinematic specification of dynamics as an informational basis for person and action perception: Expectation, gender recognition, and deceptive intention. *Journal of Experimental Psychology: General, 112*, 585-615.

Runeson, S., Juslin, P., & Olsson, H. (2000). Visual perception of dynamic properties: cue heuristic versus direct-perceptual competence. *Psychological Review, 107*, 525-555.

Runeson, S., & Vedeler, D. (1993). The indispensability of precollision kinematics in the visual perception of relative mass. *Perception & Psychophysics, 53*, 617-632.

Russell, B. (1948). *Human Knowledge*. New York: Simon and Schuster.

Salmon, W. (1994). Causality without counterfactuals. *Philosophy of Science 61*, 297-312.

Salmon, W. (1998). *Causality and explanation*. Oxford: Oxford University Press.

Schank, R. C. (1972). Conceptual dependency: A theory of natural language understanding. *Cognitive Psychology, 3,* 532-631.

Schlottman, A., & Shanks, D. R. (1992). Evidence for a distinction between judged and perceived causality. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology 44*, 321-42.

Scholl, B. J., & Tremoulet, P. D. (2000). Perceptual causality and animacy. *Trends in Cognitive Sciences, 4*, 299-309.

Scholl, B. J., & Nakayama, K. (2004). Illusory Causal Crescents: Misperceived spatial relations due to perceived causality. *Perception, 33*, 455-469.

Schwartz, D. L. (1999). Physical imagery: Kinematic versus dynamic models. *Cognitive Psychology, 38*, 433-464.

Shanks, D. R., & Dickinson, A. (1987). Associative accounts of causality judgment. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory, Vol 21* (pp. 229-261). San Diego, CA: Academic Press.

Shibatani, M. (1976). The grammar of causative constructions: A conspectus. In M. Shibatani (Ed.), *Syntax and semantics. Vol 6: The grammar of causative constructions* (pp. 1-40). New York: Academic Press.

Shultz, T. R. (1982). Rules of causal attribution. *Monographs of the Society for Research in Child Development, 47*, 1-51.

Siskind, J. M. (2000). Visual event classification via force dynamics. In *Proceedings of the American Association for Artificial Intelligence* (pp. 149-155).

Sloman, S. (2005). *Causal models: How people think about the world and its alternatives.* Oxford: Oxford University Press.

Sloman, S. (2006). The meaning of Cause and Prevent: The role of causal mechanisms. Symposium on Objects in Motion: How cause is captured in language and thought. 2006 Eastern Psychological Association, Baltimore, MD.

Sloman, S. A., & Lagnado, D. A. (2002). Counterfactual undoing in deterministic causal reasoning. In W. D. Gray & C. D. Schunn (Eds.), *Proceedings of the twenty-fourth annual conference of the Cognitive Science Society* (pp. 828-833). Hillsdale, NJ: Lawrence Erlbaum Associates.

Sloman, S. A., & Lagnado, D. A. (2005). Do we “do”? *Cognitive Science, 29*, 5-39.

Sobel, D. M., Tenenbaum, J. B., & Gopnik, A. (2004). Children's causal inferences from indirect evidence: Backwards blocking and Bayesian reasoning in preschoolers. Cognitive Science, 303-333.

Spellman, B. A. (1996). Acting as intuitive scientists: Contingency judgments are made while controlling for alternative potential causes. *Psychological Science 7,* 337-342.

Spellman, B. A., & Mandel, D. R. (1999). When possibility informs reality: Counterfactual thinking as a cue to causality. *Current Directions in Psychological Science, 8*, 120-123.

Spellman, B. A., Kincannon, A. P., & Stose, S. J. (2005). The relation between counterfactual and causal reasoning. In D. R. Mandel, D. J. Hilton, & P. Catellani (Eds.), *The psychology of counterfactual thinking*. London: Routledge Research.

Talmy, L. (1985). Force dynamics in language and thought. In W. Eilfort, P. Kroeber & K. Peterson (Eds.), *Papers from the parasession on causatives and agentivity at the 21st regional meeting, Chicago Linguistics Society* (pp. 293-337). Chicago: Chicago Linguistics Society.

Talmy, L. (1988). Force dynamics in language and cognition. *Cognitive Science, 12*, 49-100.

Tenenbaum, J. B., & Griffiths, T. L. (2001). Structure learning in human causal induction. In T. Leen, T. Dietterich & V. Tresp (Eds.), *Advances in Neural Information Processing Systems 13* (pp. 59-65). Cambridge, MA: MIT Press.

Tenenbaum, J. B., & Griffiths, T. L. (2003). Theory-based causal inference. In S. Becker, S. Thrun & K. Obermayer (Eds.), *Advances in neural information processing systems 15* (pp. 35-42). Cambridge: MIT Press.

Turnbull, W. & Slugoski, B. (1988). Conversational and linguistic processes in causal attribution. In D. Hilton (Ed.), *Contemporary science and natural explanations: Commonsense conceptions of causality* (pp. 66-93). New York: New York University Press.

Van Valin, R. D. (1990). Semantic parameters of split intransitivity. *Language, 66*, 221-260.

Verhagen, A. (2002). Interpreting usage: Construing the history of Dutch causal verbs. In M. Barlow & S. Kemmer (Eds.), *Usage-based models of language*. Stanford, CA: CSLI Publications.

Verhagen, A., & Kemmer, S. (1997). Interaction and causation: A cognitive approach to causative constructions in Modern Standard Dutch. *Journal of Pragmatics, 27*, 61-82.

Waldmann, M. R., & Hagmayer, Y. (2005). Seeing versus doing: Two modes of accessing causal knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 216-227

Wasserman, E. A., Elek, S. M., Chatlosh, D. L., & Baker, A. G. (1993). Rating causal relations: The role of probability in judgments of response-outcome contingency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 174-188.

Wells, G. L., & Gavanski, I. (1989). Mental simulation of causality. *Journal of Personality and Social Psychology, 56*, 161-169.

White, P. A. (1999). Toward a causal realist account of causal understanding. *American Journal of Psychology, 112*, 605-642.

White, P. A. (2000). Causal judgment form contingency information: The interpretation of factors common to all instances. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 1083-1102.

Wolff, P. & Gentner, D. (1996). What language might tell us about the perception of cause. In *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society* (pp. 453-458). Hillsdale, NJ: Erlbaum.

Wolff, P. (2003). Direct causation in the linguistic coding and individuation of causal events. *Cognition, 88*, 1-48.

Wolff, P., Klettke, B., Ventura, T., & Song, G. (2005). Categories of causation across cultures. In W. Ahn, G. R. L., L. B. C., A. B. Markman & P. Wolff (Eds.), *Categorization inside and* *outside of the lab: Festschrift in honor of Douglas L. Medin*. Washington, DC: American Psychological Association.

Wolff, P., & Song, G. (2003). Models of causation and the semantics of causal verbs. *Cognitive Psychology, 47*, 276-332.

Wolff, P., Song, G., & Driscoll, D. (2002). Models of causation and causal verbs. In M. Andronis, C. Ball, H. Elston & S. Neuval (Eds.), *Papers from the thirty-seventh meeting of the Chicago Linguistics Society, Main Session, Vol. 1* (pp. 607-622). Chicago: Chicago Linguistics Society.

Wolff, P., & Zettergren, M. (2002). A vector model of causal meaning. In W. D. Gray & C. D. Schunn (Eds.), *Proceedings of the twenty-fourth annual conference of the Cognitive Science Society* (pp. 944-949). Mahwah, NJ: Lawrence Erlbaum Associates.

Authors' Note

This research was supported in part by an award from the University Research Committee of Emory University. I am also very grateful to Larry Barsalou, Aron Barbey, Bianca Klettke, and Tatiana Vassilieva for their help in the data collection and analyses, as well as Aron Barbey and Larry Barsalou for their help in the construction of the animations used in Experiment 6. The materials in used in Experiment 6 were originally developed for a line of research on the brain mechanisms behind social causation in collaboration with Larry Barsalou and Aron Barbey.

Appendix A

<u>The dynamics model and Talmy's theory of force dynamics</u>

The dynamics model is based on Talmy's theory of force dynamics. However, there are several ways in which the two accounts differ, as listed below.

1. In Talmy's theory, the fundamental dimensions for distinguishing different causal concepts are 1) the intrinsic tendency of the agonist (or patient) for rest or motion, 2) the balance of strengths (i.e., the relative strength of the agonist and antagonist), and 3) the result of the force interaction. These first three dimensions account for steady-state force dynamic patterns instantiating CAUSE, PREVENT, and two types of DESPITE. In contrast, the basic dimensions in the dynamics model include two of Talmy's basic dimensions: 1) tendency of the patient and 2) occurrence of the result. However, the third dimension in the dynamics model, concordance, is not one of Talmy's basic dimensions. According to Talmy, most force dynamic interactions involve opposition between the antagonist (affector) and agonist (patient), hence the terms "antagonist" and "agonist." In the dynamics model, forces are often in concordance, as also proposed by Jackendoff (1990). These three dimensions allow for eight possible combinations of features. Six of these combinations map onto the following concepts: CAUSE (N-N-Y), PREVENT (Y-N-N), ENABLE (Y-Y-Y), DESPITE (Y-N-Y), ENABLE-NOT (N-Y-N), and DESPITE_N (N-N-N). The remaining two combinations of features (Y-Y-N & N-Y-Y) are never realized because they map onto impossible configurations of vectors. For example, the combination of features Y-Y-N implies that the patient and the affector point toward the endstate, which implies that their resultant must also point toward the endstate. If the resultant points toward the endstate, the result must occur; hence, the last feature of this combination could not be "N."

2. Talmy's first three dimensions are redundant with each other. Knowing the value of any two of the dimensions determines the value of the third. For example, if the agonist's tendency is for rest, but the agonist enters into a state of action, the balance of strengths must be that the antagonist is stronger than the agonist. As a consequence, Talmy's first three dimensions actually reduce to just two dimensions, with the intrinsic tendency of the agonist and the result being perhaps the most important. In contrast, in the dynamics model, the three main dimensions are not fully redundant with each other. This means that none of the dimensions can be removed without compromising the model's ability to distinguish different concepts.

3. Talmy's first three dimensions allow for eight (2^3) possible concepts, but the meanings of only four of them are described, specifically, CAUSE, PREVENT, and two types of DESPITE. Talmy does not explain the meanings of the remaining four possible concepts or offer an explanation for why they are rarely realized in language or experience. Indeed, three of the remaining concepts are quite odd. For example, one of these concepts represents a situation in which the agonist has a tendency for action, the antagonist is weaker than the agonist, but the agonist remains at rest. In contrast, the dynamics model is able to account for all of the possible combinations of features that can be generated from its three main dimensions.

4. In Talmy's theory, the antagonist is stronger than the agonist in CAUSE and ENABLE configurations. In the dynamics model, the antagonist is necessarily stronger only in the case of one-dimensional CAUSE configurations. In two-dimensional configurations, the antagonist need not be stronger than the agonist.

5. In Talmy's theory, all interactions are restricted to a single dimension. In the dynamics model, interactions can occur across more than one dimension. More than one dimension may be involved in causal relationships in which the cause is necessary, but not sufficient, for the effect (e.g., *Rain caused the crops to grow*).

6. In Talmy's theory, the agonist (patient) can have a tendency for either rest or motion. In the dynamics model, in contrast, tendency is defined with respect to direction toward an endstate. According to the dynamics model, when a patient is at rest or is moving away from the endstate, it is not considered to have a tendency for the endstate. Thus, the dynamics model implies two ways in which an entity might not have a tendency for the endstate while Talmy's theory implies only one way.

7. In Talmy's theory, an agonist can change state by entering into a state of motion or a state of rest. Thus, in Talmy's theory, prevention implies coming to a stop. In the dynamics model, changes of state also include changes in direction. As a consequence, in the dynamics model, prevention can involve either coming to a stop or continuing to move, but in a direction away from the endstate.

8. Talmy's force dynamics does not allow for interactions in which, for example, one antagonist causes the result while another enables it. In the dynamics model, in contrast, such situations are readily allowed through the use of multiple affector vectors. As a consequence, the

dynamics model, unlike Talmy's theory, is able to explain why certain interactions involve both causes and enabling conditions.

9. The concept of HELP is related to but different from the concepts of ALLOW or LET. Talmy (1988) accounts for the concept of LET in terms of the removal of the antagonist. Talmy's intuitions can be implemented in the dynamics model as a sequence of PREVENT configurations in which the resultant of the first configuration is used as the affector of the second (see Barbey & Wolff, 2006). For example, when we say that *Green tea prevents Alzheimer's* and *Alzheimer's prevents remembering*, this implies that *Green tea allows remembering*. According to this analysis, the concepts of ALLOW and LET are complex predicates in which something prevents a prevention, and thus allows it to occur. The dynamic models offers yet another way of representing the concepts of ALLOW and LET, specifically, by using the inverse of the **A** vector in a PREVENT configuration. The inverse of the **A** vector (**NOT-A)** would have approximately the same magnitude as **A**, but would be opposite in direction. Such an account predicts that the expression *A prevents B* implies *~A allows/lets B*. For instance, *the cat prevents the mice from playing* implies *the absence of the cat allows the mice to play*.

10. Beyond the first three basic dimensions, Talmy's theory of force dynamics includes other dimensions such as 1) addition and/or removal of a property, 2) the presence versus absence of contingent forces, and 3) increases in strength versus constancy in strength. Talmy included these dimensions to show how force dynamics might be extended to interactions beyond causal interactions. The dynamics model will also need to include additional dimensions in order to capture the range of situations examined in Talmy (1988).

Appendix B

Magnitudes of forces used in Experiment 3.

Config.	Blimp		Ice boat		Helicopter		Motor boat	
	$\|\|\mathbf{A}_w\|\|$	$\|\|\mathbf{P}_w\|\|$	$\|\|\mathbf{A}_w\|\|$	$\|\|\mathbf{P}_w\|\|$	$\|\|\mathbf{A}_w\|\|$	$\|\|\mathbf{P}_w\|\|$	$\|\|\mathbf{A}_w\|\|$	$\|\|\mathbf{P}_w\|\|$
1	1.57N	1.18N	1.7N	.90N	1.57N	1N	.591N	.394N
2	1.57N	1.18N	1.7N	.90N	1.57N	1N	.591N	.394N
3	1.18N	1.18N	.90N	.90N	1N	1N	.591N	.591N
4	1.18N	1.57N	.90N	1.7N	1N	1.57N	.394N	.590N
5	1.57N	1.18N	1.7N	.90N	1.57N	1N	.590N	.394N
6	1.18N	1.57N	.90N	1.7N	1N	1.57N	.394N	.590N

Appendix C

How momentum and energy transfer were calculated for Table 9.

Momentum is transferred when a force acting on an object changes that object's speed or direction. For example, if wind increases an object's speed, momentum from the air molecules is transferred to the object. If wind decreases an object's speed, momentum is transferred in the opposite direction. Momentum, $\boldsymbol{p}$, is given by multiplying mass, m, by velocity, $\boldsymbol{v}$, or simply $\boldsymbol{p} = m\boldsymbol{v}$. To calculate the amount of momentum transfer (i.e., impulse), the initial momentum of the object, $\boldsymbol{p}_\mathbf{i}$, is subtracted from the final momentum, $\boldsymbol{p}_\mathbf{f}$, or $\Delta\boldsymbol{p} = \boldsymbol{p}_\mathbf{f} - \boldsymbol{p}_\mathbf{i}$, In Experiments 1 and 2, the boat's mass was 1 kg. As a consequence, the change in momentum in these experiments is given by $\Delta\boldsymbol{p} = \mathbf{v}_\mathbf{f} - \mathbf{v}_\mathbf{i}$. Since the boat was accelerating before and after the fans turned on, I used the average velocity of the boat before and after the fans turned on to calculate the velocities. The average velocity is the displacement over total time, that is, $\bar{v} = \boldsymbol{d} / \Delta t$. Momentum is reported in kilograms * mass/second (kg*m/s) However, because momentum is a vector, it is also necessary to indicate the direction of the momentum vector. In Table 9, the direction of the *change* in momentum is indicated with respect to the direction of the initial velocity vector.

An object's momentum is closely linked to its kinetic energy, as reflected in their equations: whereas momentum is given by $\boldsymbol{p} = m\boldsymbol{v}$, kinetic energy is given by $KE = \frac{1}{2}m\mathbf{v}^2$. When a force increases an object's speed, it increases its kinetic energy. When a force changes on object's kinetic energy, it performs work on that object, $W=\Delta KE$. Thus, work measures energy transfer. For example, when the fans performed positive work on the boat (i.e., sped the boat up), energy (actually, air molecules) was transferred from the fans to the boat. When the fans performed negative work on the boat, energy was transferred from the boat to the air molecules. Because the boat's mass was 1kg, the energy transfer to the boats could be calculated by $W = \frac{1}{2}\mathbf{v}^2_{\text{final}} - \frac{1}{2}\mathbf{v}^2_{\text{initial}}$.